

# Studying Reionization with $\text{Ly}\alpha$ Emitters

Matthew McQuinn<sup>1\*</sup>, Lars Hernquist<sup>1</sup>, Matias Zaldarriaga<sup>1,2</sup>, Suvendra Dutta<sup>1</sup>

<sup>1</sup> *Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138*

<sup>2</sup> *Jefferson Laboratory of Physics, Harvard University, Cambridge, MA 02138*

1 February 2008

## ABSTRACT

We show that observations of high-redshift  $\text{Ly}\alpha$  emitters (LAEs) have the potential to provide definitive evidence for reionization in the near future. Using 200 Mpc radiative transfer simulations, we calculate the effect that patchy reionization has on the line profile, on the luminosity function, and, most interestingly, on the clustering of emitters for several realistic models of reionization. Reionization increases the measured clustering of emitters, and we show that this enhancement would be essentially impossible to attribute to anything other than reionization. Our results motivate looking for the signature of reionization in existing LAE data. We find that for stellar reionization scenarios the angular correlation function of the 58 LAEs in the Subaru Deep Field  $z = 6.6$  photometric sample is more consistent with a fully ionized universe (mean volume ionized fraction  $\bar{x}_i \approx 1$ ) than a universe with  $\bar{x}_i < 0.5$  at  $> 2\text{-}\sigma$  confidence level. Measurements in the next year on Subaru will increase their  $z = 6.6$  LAE sample by a factor of five and tighten these limits. If the clustering signature of reionization is detected in a LAE survey, a comparison with a Lyman-break or a  $\text{H}\alpha$  survey in the same field would confirm the reionization hypothesis. We discuss the optimal LAE survey specifications for detecting reionization, with reference to upcoming programs.

**Key words:** cosmology: theory – diffuse radiation – intergalactic medium – large-scale structure of universe – galaxies: high redshift – line: profiles

## 1 INTRODUCTION

The reionization epoch – when the hydrogen in the Universe was ionized by photons produced in the first galaxies – remains one of the least explored periods in cosmology. There is substantial uncertainty as to when reionization occurred. The latest cosmic microwave background (CMB) data show that the mean redshift of reionization was  $z_{\text{rei}} = 11 \pm 3$ , but are consistent with  $z_{\text{rei}} = 0$  at  $3\text{-}\sigma$  (Page et al. 2006), and measurements of  $\text{Ly}\alpha$  absorption in the spectra of high-redshift quasars suggest that reionization ended at  $z \approx 6$  (Becker et al. 2001; White et al. 2003; Fan et al. 2006).

Whether observations have already pinned down the redshift of bubble percolation to be  $z \approx 6$  using the  $\text{Ly}\alpha$  absorption features in the spectra of high-redshift quasars – the  $\text{Ly}\alpha$  forest – is under debate. The appearance of Gunn-Peterson absorption troughs in the  $z > 6$   $\text{Ly}\alpha$  forest may signify a change in the ionization state of the intergalactic medium (Becker et al. 2001; White et al. 2003; Fan et al. 2006). In addition, the rapid decrease of  $\text{Ly}\alpha$  forest trans-

mission at  $z > 6$  (as well as the decrease in transmission in the  $\text{Ly}\beta$  and the  $\text{Ly}\gamma$  forests), the large variance in the transmission between sight-lines, and the sizes of the proximity regions around  $z \approx 6$  QSOs may also indicate that reionization is ending (e.g., Becker et al. 2001; White et al. 2003; Fan et al. 2006; Wyithe et al. 2005; Wyithe & Loeb 2006; Mesinger & Haiman 2004; Mesinger et al. 2007).

However, owing to the resonant nature of the  $\text{Ly}\alpha$  line,  $\text{Ly}\alpha$  transmission is essentially zero if the gas is more neutral than one part in a thousand, which makes the  $\text{Ly}\alpha$  forest insensitive to the order unity fluctuations in the ionized fraction that define reionization. Because of this limitation, arguments for the end of reionization based on the forest are controversial. Becker et al. (2006) proposed that the rapid evolution of  $\text{Ly}\alpha$  transmission in the  $z \approx 6$  forest is consistent with an ionized intergalactic medium (IGM). Lidz et al. (2006) and Liu et al. (2006) showed that the observed variance in the transmission between sight-lines is consistent with the variance one expects from an ionized IGM, and Bolton & Haehnelt (2007) and Lidz et al. (2007) demonstrated that current data cannot distinguish the prox-

\* mmcquinn@cfa.harvard.edu

imity region owing to enhanced ionizing flux from the effect of an HII region.

High-redshift QSOs will probably not reveal additional information about reionization in the next few years, as all-sky surveys that observe in the near-infrared are required to push to higher redshifts. The Ly $\alpha$  forest in the spectrum of high-redshift gamma ray bursts (GRBs) holds more immediate promise, but the presence of damped Ly $\alpha$  systems within GRB host galaxies complicates the determination of the ionization state of the IGM. Despite this difficulty, through measurements of the Ly $\alpha$  and the Ly $\beta$  forests in the spectrum of a single GRB, Totani et al. (2006) derived the constraint  $\bar{x}_i > 0.4$  at  $z = 6.3$ , assuming a uniformly ionized IGM, where  $\bar{x}_i$  is the global ionized fraction of the IGM. Accounting for a patchy reionization process will weaken this limit.

Observations of high-redshift 21cm emission have the potential to put the strongest constraints on  $\bar{x}_i$  in the long term. In principle, such studies can image the neutral hydrogen in the Universe as it became ionized. However, 21cm observations must first overcome terrestrial radio broadcasts as well as galactic foregrounds that are four orders of magnitude larger than the signal (e.g., Zaldarriaga et al. 2004; McQuinn et al. 2006; Furlanetto et al. 2006). Despite these difficulties, the Mileura Widefield Array and the Low Frequency Array will start dedicated observations of this emission in the coming years (Morales & Hewitt 2004; de Vos 2004), and the 21cm Array and the Giant Meter-wave Radio Telescope are already being used to perform these measurements (Pen et al. 2005).

Studies of the CMB will never be able to provide a direct detection of reionization by imaging individual HII regions. However, measurements of CMB anisotropies can improve constraints on the global properties of reionization. A more precise measurement of CMB polarization anisotropies at low multipoles has the potential to provide some information about the duration of reionization in addition to tighter constraints on the average redshift of reionization (e.g., Keating & Miller 2006). The *Planck* mission will measure the large-scale E-mode polarization anisotropies to nearly the cosmic-variance limit in the next few years. In addition, a sizable fraction of arc-minute-scale CMB anisotropies are imprinted during patchy reionization (Santos et al. 2003; Zahn et al. 2005; McQuinn et al. 2005), and the Atacama Cosmology Telescope and the South Pole Telescope are starting to observe these anisotropies.<sup>1</sup> These observations have the potential to provide information about the duration of reionization and about the sizes of the HII regions.

In this paper, we demonstrate that Ly $\alpha$  emitter (LAE) surveys will likely offer the first irrefutable proof as to precisely when reionization occurred, at least if the Universe is significantly neutral for  $z \lesssim 7$  as the high-redshift quasars suggest. We show that the  $z = 6.6$  LAE sample from Subaru can already put limits on  $\bar{x}_i$  with just 58 photometrically confirmed LAEs. Surveys on existing telescopes will also put constraints on the neutral fraction at higher redshifts, and the James Webb Space Telescope (JWST) – with the appro-

priate observing strategy – will effectively be able to image HII regions during reionization.

There are several surveys that have targeted or are targeting LAEs at epochs when the Universe may have been significantly neutral (e.g., Kodaira et al. 2003; Rhoads et al. 2004; Kashikawa et al. 2006; Cuby et al. 2007; Willis & Courbin 2005; Stark et al. 2007), and other programs will begin taking data soon (e.g., Casali et al. 2006; McPherson et al. 2006; Horton et al. 2004). These surveys take advantage of the fact that a galaxy can produce a sizable fraction of its flux in the Ly $\alpha$  line (Partridge & Peebles 1967). The transmission of the Ly $\alpha$  line from a galaxy is decreased if the neighborhood of an emitter is largely neutral, and this modulation can be used to probe the epoch of reionization (Miralda-Escude 1998; Madau & Rees 2000; Haiman 2002; Santos 2004; Furlanetto et al. 2004c; Furlanetto et al. 2006; Malhotra & Rhoads 2006a).

Most previous theoretical studies of high-redshift LAEs have focused on using the evolution in the LAE luminosity function or in the average Ly $\alpha$  line profile to constrain reionization. The first predictions for the impact of reionization on the LAE luminosity function and line profile were based on the assumption that each emitter sits in its own HII region until the final bubble percolation stage, at which time all of the HII regions quickly merge and reionization is completed (Santos 2004; Haiman & Cen 2005). In reality, the HII regions can be much larger throughout reionization than this simple model predicts owing to the clustering of high-redshift sources (e.g., Sokasian et al. 2003; Furlanetto et al. 2004a,b). When the Universe has  $\bar{x}_i = 0.5$ , typically a LAE will be in an HII region created by thousands of galaxies. Because of this clustering, both the evolution of the LAE luminosity function and the change in the shape of the average line profile of an emitter are a weaker function of the neutral fraction than predicted by models based on a single source per HII region (Furlanetto et al. 2004c; Furlanetto et al. 2006; McQuinn et al. 2006).

The measured  $z = 6.5$  LAE luminosity function has already been used to constrain  $\bar{x}_i$ . Malhotra & Rhoads (2004) argued that the data is consistent with no evolution in the luminosity function between  $z = 5.7$  and  $z = 6.5$ , and this lack of evolution allowed Malhotra & Rhoads (2006b) to derive the limit  $\bar{x}_i \gtrsim 0.2$  at  $z = 6.5$ . However, in the newest and largest LAE survey at  $z > 6$ , Kashikawa et al. (2006) finds significant evolution between the  $z = 5.7$  and the  $z = 6.6$  luminosity functions for luminosities  $> 2 \times 10^{42}$  erg s<sup>-1</sup>, and they suggest that this evolution could be evidence for reionization. Furthermore, Iye et al. (2007) finds that the number density of LAEs with luminosities above  $1 \times 10^{43}$  erg s<sup>-1</sup> at  $z = 7$  is 20% – 40% smaller than this number density at  $z = 6.6$ .

In addition to the luminosity function and the Ly $\alpha$  line profile, the HII bubbles will influence the *measured* clustering of LAEs (Furlanetto et al. 2006; McQuinn et al. 2006). Detecting reionization through its impact on observed clustering will be the most fool-proof method to identify pockets of neutral hydrogen in the IGM with LAEs, since astrophysical uncertainties and observational systematics cannot induce large-scale correlations similar to  $\sim 10$  Mpc HII regions during reionization. We show that a LAE survey using clustering measurements can confirm whether reionization is occurring at  $z = 6.6$  in the immediate future and that

<sup>1</sup> <http://www.hep.upenn.edu/act/act.html>,  
<http://astro.uchicago.edu/spt/>

upcoming surveys will place constraints on reionization at higher redshifts.

This paper is organized as follows. In Section 2, we describe the simulations of reionization that we employ as well as the post-processing calculations used to generate mock Ly $\alpha$  surveys from the simulation outputs. These mock LAE surveys are featured in Section 3. The effect of reionization on the luminosity function is discussed in Section 4 (and on the line profile in Appendix A). However, the primary focus of this work is to investigate the impact that reionization has on the measured clustering properties of LAEs and to quantify the detectability of reionization-induced clustering in present surveys (Section 5 and, for void statistics, Appendix B). We conclude in Section 6 with a discussion of the capabilities of upcoming LAE surveys to constrain reionization.

## 2 CALCULATIONS

### 2.1 Simulations

We use two  $1024^3$  N-body simulations generated with the TreePM code L-Gadget-2 (Springel 2005) to model the density field, one in a box of size 94 Mpc with outputs every 50 Myr and the other in a box of size 186 Mpc with outputs every 20 Myr. A Friends-of-Friends algorithm with a linking length of 0.2 times the mean inter-particle spacing is used to identify halos. We employ the higher-resolution 94 Mpc box to study line properties. This box is run using a  $\Lambda$ CDM cosmology with  $n_s = 1$ ,  $\sigma_8 = 0.9$ ,  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ ,  $\Omega_b = 0.04$ , and  $h = 0.7$ , consistent with the first year WMAP results (Spergel et al. 2003). We use the 186 Mpc box to study the luminosity function and clustering properties of emitters. This volume provides a larger sample of the structures that are present during reionization than does the smaller box. The 186 Mpc box is run with a  $\Lambda$ CDM cosmology updated to be more consistent with the latest WMAP results (Spergel et al. 2006) with  $n_s = 1$ ,  $\sigma_8 = 0.8$ ,  $\Omega_m = 0.27$ ,  $\Omega_\Lambda = 0.73$ ,  $\Omega_b = 0.046$ , and  $h = 0.7$ . The assumptions about the cosmology do not significantly affect the morphology of reionization if we compare at fixed  $\bar{x}_i$  (McQuinn et al. 2006).

The halo mass function measured from the N-body simulations matches the Sheth & Tormen (2002) mass function for groups with 32 or more particles. Thirty-two particle groups correspond to a halo mass of  $m = 1 \times 10^9 M_\odot$  in the 94 Mpc box and  $m = 8 \times 10^9 M_\odot$  in the 186 Mpc box. We would like to resolve halos down to the atomic hydrogen cooling mass  $m_{\text{cool}}$  – halos with virial temperature of  $10^4$  K – which corresponds to the minimum mass galaxy that can form stars ( $m_{\text{cool}} \approx 10^8 M_\odot$  at  $z = 8$ ). We add unresolved halos (halos that would be comprised of fewer than 32 particles) into the simulation using a Press-Schechter merger tree. This algorithm is similar to the PThalo method for generating mock catalogs of galaxies (Sheth & Lemson 1999), and accounts for both density and Poisson fluctuations in the abundance of halos. We demonstrate in McQuinn et al. (2006) that this method reproduces well the power spectrum and mass function of dark matter halos seen in simulations.

We use an improved version of the Sokasian et al. (2001, 2003, 2004) and McQuinn et al. (2006) radiative transfer

code. We have refined the algorithm to group sources more efficiently for our work, and this improvement is discussed in Appendix D. Our approach is optimized to simulate reionization, making several justified simplifications to speed up the computation. The code takes the gridded density field (generated from the N-body simulation assuming that the baryons trace the dark matter) and a list of the ionizing sources as input, and it casts rays from each source to compute the ionization field. We assume that the sources have a soft UV spectrum that scales as  $\nu^{-4}$  (consistent with a POPII IMF), which is used to calculate the photo-ionization state of the gas. The radiative transfer code assumes perfectly sharp HII fronts, tracking the front position at sub-grid scales.<sup>2</sup> Sharp fronts are an excellent approximation for sources with a soft UV spectrum, where the mean free path for ionizing photons is kiloparsecs, substantially smaller than the cell size in the radiative transfer calculations.

The radiative transfer calculations presented in this paper were performed on a  $256^3$  grid for the 94 Mpc box and on a  $512^3$  grid for the 186 Mpc box. The radiative transfer calculations took between three days and two weeks on a 2.2 GHz AMD Opteron processor.

### 2.2 Ly $\alpha$ Line Calculation

The shape and energetics of a Ly $\alpha$  line from a galaxy are determined by complex processes, many of which cannot be modeled accurately. The Ly $\alpha$  emission depends on the amount of dust in the galaxy, the velocity profile of the gas in the halo, and the fraction of ionizing photons that escape from the galaxy and influence the photo-ionization state of the nearby IGM (Hansen & Oh 2006; Dijkstra et al. 2007; Santos et al. 2003; Tapken et al. 2007; Tasitsiomi 2006). Statistics that can be used to isolate the effect that reionization has on the emitters from all of the complicated, uncertain astrophysics are essential to probe reionization with LAEs. In order to construct mock surveys from which we can measure various statistics, we must make simplifying assumptions about the Ly $\alpha$  emission that escapes from the vicinity of a galaxy.

To calculate the emitter line profile and its transmission we employ two schemes, a detailed, expensive method and a fast, relatively inexpensive technique.

*Method 1:* We assume a Gaussian intrinsic line profile – the line profile that escapes from the vicinity of the galaxy – with its width set by the circular velocity at the virial radius of the emitter host halo ( $\sigma_\nu = \nu_\alpha v_{\text{vir}}/c$ ). This intrinsic line profile is what is commonly assumed in the literature (e.g., Santos 2004 and Dijkstra et al. 2007), but the outgoing line can be significantly more complicated (Hansen & Oh 2006; Dijkstra et al. 2006; Tapken et al. 2007; Tasitsiomi 2006). See the discussion in Appendix A for the physical motivation for the assumed line profile. In the absence of a galactic wind, we assume the intrinsic line is centered at Ly $\alpha$ . Given the intrinsic line from an emitter, we compute the absorption along a skewer through the simulation box, using the density, velocity, and photo-ionization state of the gas. The radiative transfer simulations do not compute the temperature of the

<sup>2</sup> This is not true for self-shielded regions, which can remain neutral behind the front.

gas self-consistently. For simplicity, we take the gas in all regions to be at  $10^4 K$ . The value of the gas temperature and the amplitude of the sub-grid-scale density fluctuations do not influence the amount of damping wing absorption – the effect that is most relevant to this study.

*Method 2:* First, we set the residual neutral fraction within each HII region to zero. The residual neutral gas primarily influences frequencies blueward of  $\text{Ly}\alpha$ , which this method assumes are fully absorbed. Next, we calculate the damping wing optical depth to absorption  $\tau_\alpha$  along a ray through the IGM, ignoring the effect of peculiar velocities. We do this for only a single frequency that starts off at  $\nu' = \nu_0(1 - v_w/c)$  for each source, where  $\nu_0$  is the line center and  $v_w$  is the wind velocity (which is typically taken to be zero).<sup>3</sup> We then set  $L_{\text{obs},\alpha} = \mathcal{T}_\alpha L_\alpha \exp[-\tau_\alpha(\nu')]$ , where  $\mathcal{T}_\alpha$  accounts for resonant absorption and is typically between 0.1 and 0.5 in the absence of outflows (Santos et al. 2003 and Appendix A). The exact value of  $\mathcal{T}_\alpha$  or its scaling with halo mass is not important for our present study since  $\mathcal{T}_\alpha$  is degenerate with the luminosity of the emitter prior to absorption in the IGM,  $L_\alpha$ . This method is in contrast to Method 1 in which for each source we calculate  $\tau_\alpha$  at many different frequencies.

It turns out that both methods for computing the line profile result in very similar results. (See Figure A2 for a comparison of the two methods.) The agreement between the two methods occurs because for observed emitters the virial velocity of the halo, which sets the intrinsic width of the line in our model, is typically much smaller than the relative velocity between the LAE and the HII front. The agreement between the two methods justifies the use of the cheaper method, Method 2, for computations in which the shape of the line profile is not required.

For reference, the radius of a bubble that produces a damping wing absorption cross section of unity at source-frame frequency  $\nu \sim \nu_0$  is (Loeb & Rybicki 1999)

$$R_b \approx 1.1 \bar{x}_H \left( \frac{\Omega_b}{0.046} \frac{0.27}{\Omega_m} \right) \text{pMpc} + \frac{\nu - \nu_0}{\nu_0} \frac{c}{H(z_g)}, \quad (1)$$

where pMpc denotes proper Mpc,  $\bar{x}_H$  is the neutral fraction outside of the bubble, and  $z_g$  is the redshift of a LAE in the center of the bubble. Equation (1) is correct in detail only for a homogeneous  $x_H$ . In practice, the IGM is far from uniform and the effective  $\bar{x}_H$  along each line of sight is not equal to the globally averaged  $\bar{x}_H$ . Also note that  $\tau_\alpha(\nu) \approx 900 \text{ km s}^{-1} \bar{x}_H [(1 + z_g)/8]^{3/2} [H(z_g)R_b - c(\nu - \nu_0)/\nu_0]^{-1}$  – the optical depth scales inversely with the bubble radius. (See Miralda-Escude 1998 for the exact expression for  $\tau_\alpha$  as a function of the bubble size.) If a LAE produces its own HII bubble then the size of this bubble is

$$R_b = 0.15 \left( \frac{\dot{S}}{M_\odot \text{ yr}^{-1}} \frac{t_{\text{age}}}{10^8 \text{ yr}} \frac{f_{\text{esc}}}{0.1} \right)^{\frac{1}{3}} \left( \frac{8}{1 + z} \right) \text{pMpc}, \quad (2)$$

where  $\dot{S}$  is the star formation rate, which is estimated to be approximately  $1\text{--}10 M_\odot \text{ yr}^{-1}$  for observed emitters (Taniguchi et al. 2005), and  $t_{\text{age}}$  is the lifetime of the emitter. Equation (2) does not account for recombinations and

<sup>3</sup> Note that the actual wind velocity is probably closer to  $0.5 v_w$  such that  $v_w$  characterizes the shift of the  $\text{Ly}\alpha$  line rather than the wind velocity.

assumes a Salpeter IMF with  $1.5 \times 10^{53} \dot{S}/(M_\odot \text{ yr}^{-1})$  ionizing photons  $\text{s}^{-1}$  (Hui et al. 2002). Note that typically  $\tau_\alpha(\nu_0) > 1$  for an emitter that sits in an HII region created entirely by itself. In CDM theory, high redshift galaxies are very clustered such that the situation of a single source creating an HII region is a rarity.

### 2.3 Simple Model for Intrinsic Distribution of LAEs

Given the methods discussed in Section 2.2 for calculating the line profile, we also need an algorithm that tells us where the LAEs are located to generate mock surveys. For the analysis in this paper, we adopt a simple model to populate the halos in the simulation with LAEs. We describe a mock survey of LAEs with three parameters:

- $m_{\text{min}}$  – minimum mass halo that hosts an observed LAE
- $f_E$  – fraction of halos that host LAEs
- $F_c$  – fraction of the LAE sample contaminated by lower-redshift interlopers

We assume that each halo in the simulation box has at most one LAE, which is reasonable at  $z > 6$  since the gas cooling time in all halos is shorter than a Hubble time. We also assume that  $f_E$  is not a function of halo mass.<sup>4</sup> This model is similar to those for LAEs used in Dijkstra et al. (2006) and in Stark et al. (2007). Both Dijkstra et al. (2006) and Stark et al. (2007) find that this simple parameterization is able to provide adequate fits to the  $z = 5.7$  and  $z = 6.6$  luminosity functions even when fixing  $F_c$ .

In the context of this model, we can write expressions for the intrinsic number density of observed emitter candidates  $\bar{n}_E = f_E(1 + F_c)\bar{n}_h(m_{\text{min}})$  as well as the intrinsic 3-D  $\bar{n}_E$ -weighted power spectrum of emitter candidates  $P_{\text{int},E}(\mathbf{k}) = (1 - F_c)^2 P_{\text{hh}}(\mathbf{k}, m_{\text{min}}) + 1/\bar{n}_E$ . Here, we use “intrinsic” to mean the value of these functions when the effect of damping wing absorption from the HII regions is not included. Note that  $\bar{n}_h(m_{\text{min}})$  is the number density of halos with  $m > m_{\text{min}}$ ,  $P_{\text{hh}}(\mathbf{k}, m_{\text{min}})$  is the power spectrum of halos that have  $m > m_{\text{min}}$  (with the shot noise component subtracted off), and this expression for  $P_{\text{int},E}$  assumes that there is no clustering of the foreground interlopers.<sup>5</sup>

We ignore peculiar velocities in our analysis. Peculiar velocities most significantly affect Fourier modes oriented along the line of sight. In narrow-band LAE surveys, the majority of the signal is instead contributed by transverse,

<sup>4</sup> Even though this assumption might not be true in detail, the majority of observed emitters reside in halos of roughly the same mass in our model because current observations probe the tail of the luminosity function.

<sup>5</sup> For narrow band LAE surveys, foreground interlopers come from a limited number of redshifts because they must have a strong emission line that falls within the narrow frequency band. As a consequence of this redshift selection, the clustering of foreground interlopers is more significant than if the interlopers came from all intervening redshifts. However, unless the contamination fraction of a LAE survey is unusually high, the extra clustering arising from the foreground interlopers should be much smaller than the contribution from the LAEs, justifying the no-clustering approximation.

long wavelength modes. Moreover, in linear theory the peculiar velocity terms that contribute to  $P_{\text{int,E}}$  are suppressed by a factor  $\sim 1/b$ , where  $b$  is the  $\bar{n}_E$ -weighted linear bias of emitters. Based on the abundance of observed  $z = 6.6$  LAEs, a reasonable estimate is  $b \approx 7$  in present surveys.

This model gives the number density and clustering of emitters with  $m > m_{\text{min}}$ . However, surveys do not directly infer the mass of an emitter but instead measure its Ly $\alpha$  luminosity. In the absence of dust, we can map the ionizing luminosity from a source in our simulations to its luminosity in Ly $\alpha$  photons via (Osterbrock 1989)

$$L_\alpha = 0.67 E_\alpha (1 - f_{\text{esc}}) \dot{N}_{\text{ion}}, \quad (3)$$

where  $E_\alpha = 10.2$  eV,  $f_{\text{esc}}$  is the fraction of ionizing photons that escape into the IGM, and  $\dot{N}_{\text{ion}}$  is the production rate of ionizing photons. Assuming that equation (3) holds, that  $f_{\text{esc}} = 0.02$ , and that  $L_{\text{obs,E}} = 0.5 L_\alpha$  (where the factor 0.5 roughly accounts for resonant absorption in the IGM), the observed Ly $\alpha$  luminosity of the sources in our fiducial reionization model [model (i) in the next section] is  $L_{\text{obs,E}} = 5 \times 10^{42} m / (10^{11} M_\odot) \text{ erg s}^{-1}$ . The emitters that have been observed at  $z = 6.6$  have  $L_{\text{obs,E}} = (2 - 20) \times 10^{42} \text{ erg s}^{-1}$  (Taniguchi et al. 2005). Lowering or raising  $f_{\text{esc}}$  – which is currently unconstrained – can decrease or increase  $L_{\text{obs,E}}$  for a halo of fixed  $m$ .

## 2.4 Three Reionization Models

We consider three models for the reionization history. These models bracket the possible range of morphologies for reionization by POPII-like stars (as shown in McQuinn et al. 2006):

(i) All halos above  $m_{\text{cool}}$  contribute ionizing photons at a rate that is proportional to their mass  $m$ . The ionizing luminosity of source halos is  $\dot{N}_{\text{ion}}(m) = 3 \times 10^{51} [m / (10^{10} M_\odot)]$  ionizing photons  $\text{s}^{-1}$ . The scaling  $\dot{N}_{\text{ion}} \sim m$  assumes that the star formation rate is proportional to the amount of gas within a galaxy. The normalization of  $\dot{N}_{\text{ion}}$  is chosen such that reionization ends at  $z = 7$ . Given the uncertainty in  $f_{\text{esc}}$  and the star formation rate in high-redshift galaxies, the range of possible normalizations is vast. Fortunately, the morphology of reionization depends very weakly on the normalization of  $\dot{N}_{\text{ion}}$ , as shown in McQuinn et al. (2006).

(ii) Halos more massive than  $m_{\text{cool}}$  contribute to the production of ionizing photons, with the ionizing luminosity of the sources scaling as halo mass to the 5/3 power (i.e., more massive halos dominate the production of ionizing photons than in model (i)). This scaling is chosen to match the relationship between star formation efficiency and galaxy mass that is observed in low-redshift dwarf galaxies (Kauffmann et al. 2003) as well as the star formation rate found in theoretical studies that include supernova feedback (Dekel & Woo 2003; Springel & Hernquist 2003). The total budget of ionizing photons released is calibrated such that reionization ends at  $z \approx 7$  as in model (i).

(iii) Absorption by minihalos shapes the morphology of reionization. We use the same source parametrization as in model (i) except that the sources are twice as luminous so that reionization ends at  $z \approx 7$ . While minihalos do not contribute ionizing photons in this model, they do act as

photon sinks. All minihalos with  $m > 10^5 M_\odot$  absorb incident ionizing photons out to their virial radius until they are photo-evaporated. This absorption cross section is larger than the cross section found in radiative-hydrodynamic simulations of minihalo evaporation (Iliev et al. 2005), suggesting that this model overestimates the impact of minihalo absorption. The photo-evaporation timescale is roughly the sound-crossing timescale of a halo.<sup>6</sup>

Figure 1 displays slices through simulations using models (i), (ii), and (iii). The white regions are ionized and the black are neutral. Model (ii) results in the largest HII regions because it has the most biased sources, whereas model (iii) produces the smallest bubbles, with the maximum bubble radius restricted to be roughly the mean free path for ionizing photons to intersect a minihalo.

One piece of physics that is missing from these three reionization models and that has not been quantified in simulations of reionization is the effect of a duty cycle for the ionizing sources on the morphology. It is probable that the galaxies form massive stars and contribute ionizing photons only during certain periods. In Appendix C, we demonstrate that the duty cycle of the sources does not affect the morphology of the ionized regions for most realistic reionization models.

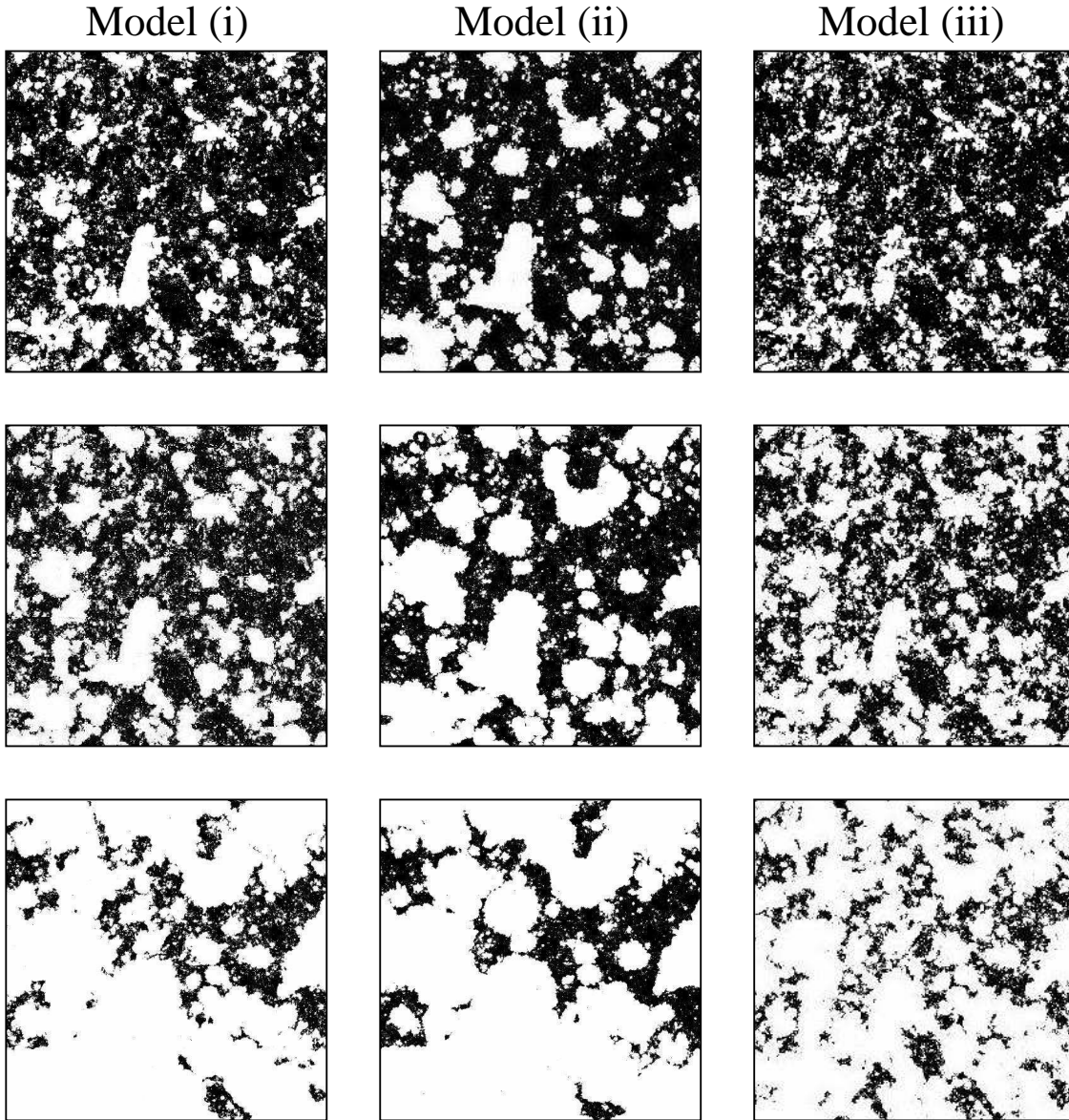
## 3 LAE MAPS

Figure 2 shows mock LAE surveys created using snapshots from the simulation of model (i) in the 94 Mpc box and with a depth of 130 Å or approximately 35 Mpc. The dimensions of these mock surveys are roughly the same as the  $z = 6.6$  Subaru Deep Field (Taniguchi et al. 2005; Kashikawa et al. 2006). Each panel would subtend 0.6 degrees or roughly the solid angle of the moon. These LAE maps are generated using Method 2 in Section 2.2 and assuming that 25% of halos host emitters<sup>7</sup> (i.e.,  $f_E = 0.25$ ) and that only those Ly $\alpha$  emitting halos can be observed that have  $m \exp(-\tau_\alpha(\nu_0)) > 7 \times 10^{10} M_\odot$ . This model corresponds to setting  $m_{\text{min}} = 7 \times 10^{10} M_\odot$ . Note that  $L_{\text{int,E}} \propto m$  in model (i).

The top panels in Figure 2 are the projected ionization maps for  $\bar{x}_i = 0.3$  (*left panel*), 0.5 (*middle panel*), and 0.7 (*right panel*) at redshifts  $z = 8.2$ , 7.7, and 7.3, respectively. The projected map is completely ionized in white areas and neutral in black ones. The middle row of panels shows the intrinsic distribution of emitters (or what would be observed if  $\bar{x}_i \approx 1$ ), and the bottom row of panels shows the observed distribution. (Compare the bottom panels with their corresponding top panels to see the reionization-induced modulation.) The “Intrinsic”  $\bar{n}_E$  in Figure 2 is a few times higher

<sup>6</sup> We use the fitting formula in Iliev et al. (2005) for the evaporation timescale, which parameterizes the evaporation timescale as a function of redshift, halo mass, and incident ionizing flux.

<sup>7</sup> Twenty-five percent is chosen to match the fraction of Lyman-break galaxies at  $z = 3$  that meet the selection criteria to be detected in a narrow band survey as a LAE (Shapley et al. 2003), although it is probable that Lyman break galaxies have some duty cycle as well and that the fraction of galaxies that emit in Ly $\alpha$  evolves with redshift.

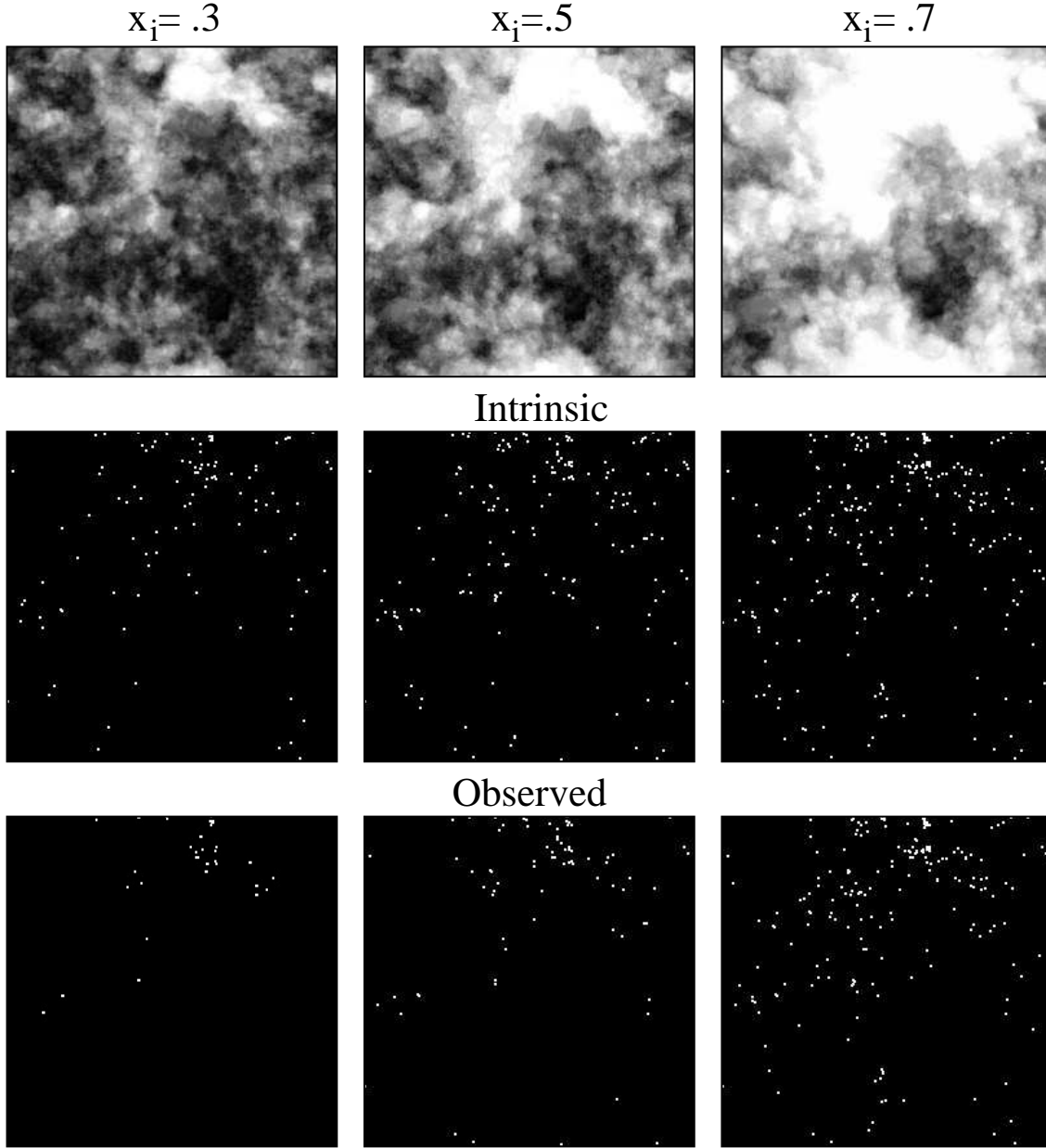


**Figure 1.** Slices through the middle of the 186 Mpc simulation box for the three models. The top row is at  $\bar{x}_i = 0.3$ , the middle is at  $\bar{x}_i = 0.5$ , and the bottom is at  $\bar{x}_i = 0.8$ . Model (iii), in which minihalos limit the photon mean free path, has the smallest bubbles, whereas model (ii), in which the sources are the most biased, has the largest.

then the  $\bar{n}_E$  of the  $z = 6.6$  Subaru Deep Field (SDF) photometric sample (Taniguchi et al. 2005). The “Observed”  $\bar{n}_E$  when  $\bar{x}_i = 0.5$  in Figure 2 is comparable to the  $\bar{n}_E$  of the  $z = 6.6$  SDF photometric sample. Note the higher degree of clustering in the “Observed” panels relative to the “Intrinsic” panels. We investigate the detectability of this clustering in Section 5.

Figure 3 is the same as Figure 2, but for a futuristic survey that is sensitive to  $\text{Ly}\alpha$  emitting halos with  $m \exp(-\tau_\alpha(\nu_0)) > 1 \times 10^{10} M_\odot$ . There are approximately 2000 sources in the “Intrinsic” panels, but there can be significantly fewer in the “Observed” panels. The large-scale modulation of emitters by the bubbles in Figure 3 allows observations to image the neutral holes in the map with the LAEs. If the  $\text{Ly}\alpha$  luminosity scales roughly with  $m$ , as it does for the sources in these calculations, the mock sur-

vey depicted in Figure 3 would entail a  $\approx 50$  times longer observation than that of Figure 2 (assuming that the observation is photon-limited). A space mission with a wide field of view (FoV) camera could provide such a deep survey in a much shorter observation. The James Webb Space Telescope (JWST), set to be launched in 2013, will be able to easily provide such a deep image. However, the current high-redshift program for JWST specifies a deep survey that will cover less than one-tenth of the solid angle subtended by a single panel in Figure 3 (Gardner et al. 2006). With such a small field, it will be difficult for JWST to probe the HII regions during reionization. The small square in the lower left-hand panel represents the FoV of JWST.



**Figure 2.** Top panels show the projection of  $\bar{x}_i$  in the survey volume. In the white regions the projection is fully ionized and in black it is neutral. The left, middle, and right panels are for  $z = 8.2$  ( $\bar{x}_i = 0.3$ ),  $z = 7.7$  ( $\bar{x}_i = 0.5$ ), and  $z = 7.3$  ( $\bar{x}_i = 0.7$ ). The middle and bottom rows are the intrinsic and observed LAE maps, respectively, for  $f_E = 0.25$  and assuming that we can observe unobscured emitters with  $m \exp(-\tau_\alpha(\nu_0)) > 7 \times 10^{10} M_\odot$ . (Note that  $L_{\text{int,E}} \propto m$ .) The observed distribution of emitters is modulated by the location of the HII regions (compare bottom panels with corresponding top panels). Each panel is 94 Mpc across (or 0.6 degrees on the sky), roughly the area of the current Subaru Deep Field (SDF) at  $z = 6.6$  (Kashikawa et al. 2006). The depth of each panel is  $\Delta\lambda = 130 \text{ \AA}$ , which matches the FWHM of the Subaru 9210  $\text{\AA}$  narrow band filter. The number densities of LAEs for the panels in the middle row are a few times larger than the number density in the SDF photometric sample of  $z = 6.6$  LAEs.

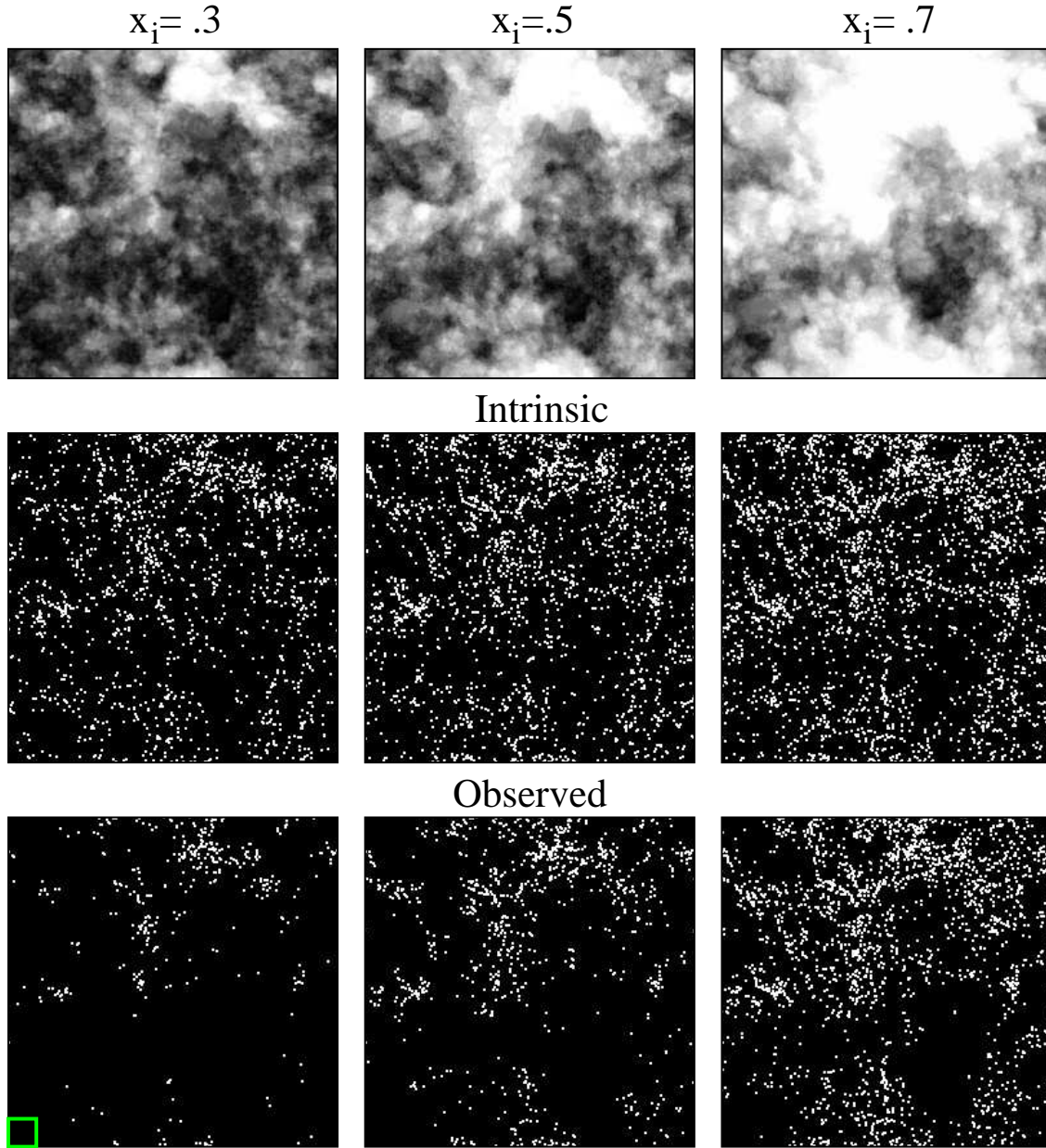
#### 4 THE LUMINOSITY FUNCTION

The luminosity function of LAEs depends sensitively on the morphology of HII regions during reionization. If all the HII regions are smaller than 1 pMpc (such that  $\tau_\alpha(\nu_0) > 1$ ), only a small fraction of LAEs would be observed compared to the number that would be observed if the Universe were fully ionized. In fact, even if the bubbles are a few times larger than 1 pMpc, many emitters will be obscured. This is because the steep, decreasing nature of the luminosity func-

tion means that the majority of emitters have luminosities that are within a factor 2 of the detection threshold, requiring for many LAEs that  $\tau_\alpha(\nu_0)$  be significantly less than unity to be observed.

As reionization proceeds, larger HII regions will form, allowing more LAEs to appear out of the dark. An extremely rapid decrease in  $\bar{n}_E$  would be difficult to attribute to evolution in the intrinsic properties of the LAEs rather than to reionization. Since the LAE luminosity function has been





**Figure 3.** Same as Figure 2, but for a futuristic LAE survey that can detect halos down to  $m \exp(-\tau_\alpha(\nu_0)) > 1 \times 10^{10} M_\odot$  (note  $L_{\text{int,E}} \propto m$ ) and assuming  $f_E = 0.25$ . The large-scale modulation of LAE by the HII bubbles is clearly apparent in this survey. The square in the lower left-hand panel represents the  $3' \times 3'$  FOV of JWST drawn to scale.

measured at  $z = 6.6$  (and tentatively at  $z = 7$ ) and will be constrained at even higher redshifts in the coming years, it is important to understand the signature of reionization in the luminosity function.

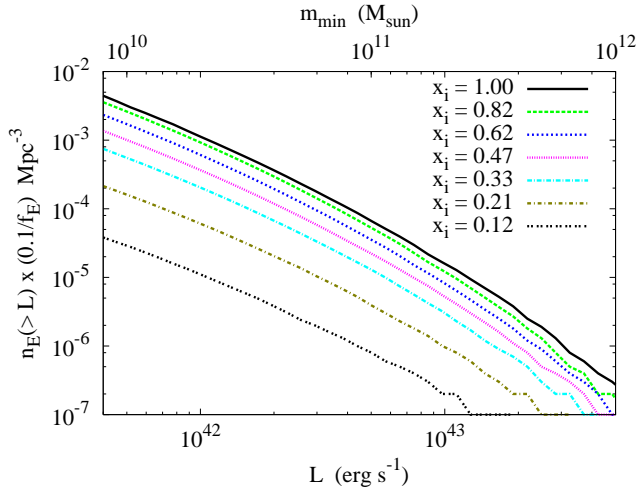
The current data on LAE luminosity functions may indicate that reionization is happening at  $z = 6.6$ . Kashikawa et al. (2006) finds that there is a suppression in the bright end of their measured luminosity function at  $z = 6.6$  relative to that at  $z = 5.7$  at  $2\text{-}\sigma$  significance. See the thick solid curve in Figure 5 for the ratio of the  $z = 5.7$  and  $z = 6.6$  best-fit LAE luminosity functions along with an estimate for the  $1\text{-}\sigma$  shot noise errors on

this ratio (Shimasaku et al. 2006; Kashikawa et al. 2006).<sup>8</sup> Kashikawa et al. (2006) propose that the suppression of the high luminosity end may imply a change in the ionization state of the IGM. In addition, Iye et al. (2007) finds an additional factor of few suppression in the luminosity function at  $z = 7$  for emitters with  $L_{\text{obs,E}} > 1 \times 10^{43} \text{ erg s}^{-1}$ .

Dijkstra et al. (2006) suggest that there is a more mundane explanation for this suppression – the evolution of the halo mass function. Employing a similar model for LAEs

<sup>8</sup> The cosmic variance errors are highly correlated between different luminosity bins, and a  $1\text{-}\sigma$  cosmic fluctuation will raise or suppress this ratio by  $\approx 50\%$  for an ionized universe (see Section 6).



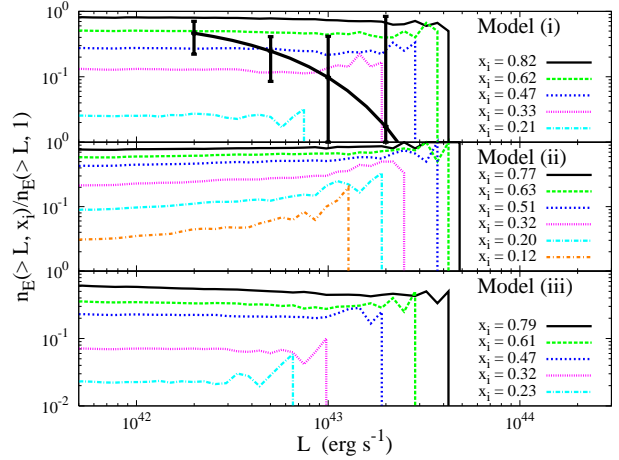


**Figure 4.** Number density of LAEs with  $L_{\text{obs,E}} > L$  at  $z = 6.6$  and for several different volume-averaged ionization fractions. These curves are calculated from a simulation of model (i) in the 186 Mpc box. The top axis shows the number density with  $m \exp(-\tau_\alpha(\nu_0)) > m_{\text{min}}$ . The mapping between  $m$  and  $L_{\text{obs,E}}$  is discussed in the text.

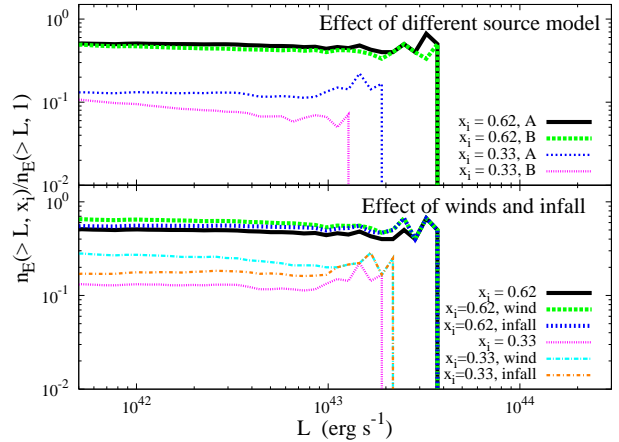
to ours, they argue that the cosmological evolution of the halo mass function between  $z = 5.7$  and  $z = 6.6$  can account for the observed suppression and that it is unnecessary to invoke reionization. However, if Dijkstra et al. (2006) were to include dispersion in the luminosity for a given halo mass, the high mass end of the luminosity function would be less dependent on the evolution of the halo mass function, possibly altering this conclusion. Also, the model in Dijkstra et al. (2006) favors  $f_E \sim 1$  in order to fit the data at  $z = 6.6$  whereas they find that  $f_E \ll 1$  provides a better fit at  $z = 5.7$ . A similar trend for the evolution of  $f_E$  is also found in Stark et al. (2007), even though such dramatic evolution in the intrinsic properties of emitters is not expected.

For most of the calculations in this section, we assume the simple model for LAEs discussed in Section 2.3. Halos that are active LAEs have  $L_{\text{int,E}}(m) = 5 \times 10^{42} (f_{\text{esc}}/0.02)^{-2} [m/(10^{11} M_\odot)] \text{ erg s}^{-1}$  in model (i). We set  $f_{\text{esc}} = 0.02$  to roughly match the observed abundance of LAEs for  $f_E = 0.1$  and  $m_{\text{min}} = 5 \times 10^{10} M_\odot$ . For the other two models, we take  $L_{\text{int,E}}(m)$  to be the same function as in model (i). To be consistent with equation (3) and the  $\dot{N}_{\text{ion}}$  used in models (ii) and (iii), fixing  $L_{\text{int,E}}(m)$  requires adjusting  $f_{\text{esc}}$  slightly.

The curves in Figure 4 represent the observed number density of LAEs with luminosities that satisfy  $L_{\text{obs,E}} > L$ . These curves are calculated from the 186 Mpc simulation of model (i) at  $z = 6.6$ .<sup>9</sup> During reionization, the luminosity function is suppressed by a factor that does not depend strongly on  $L$ . Figure 5 plots the ratio of the luminosity function at various  $\bar{x}_i$  for the three models. As with model



**Figure 5.** Ratio of the number density of LAEs with  $L_{\text{obs,E}} > L$  for different  $\bar{x}_i$ ,  $n_E(>L, \bar{x}_i)$ , to the number density for  $\bar{x}_i \approx 1$ ,  $n_E(>L, 1)$ . The thick solid curve with the 1- $\sigma$  errors in the top panel is the ratio of the  $z = 6.6$  and  $z = 5.7$  SDF luminosity functions. The effect of reionization in all models is approximately a uniform suppression of the luminosity function that is independent of luminosity. Model (ii) has the largest bubbles such that the luminosity function is the least suppressed at fixed  $\bar{x}_i$ , whereas model (iii) has the smallest bubbles such that the luminosity function is the most suppressed.

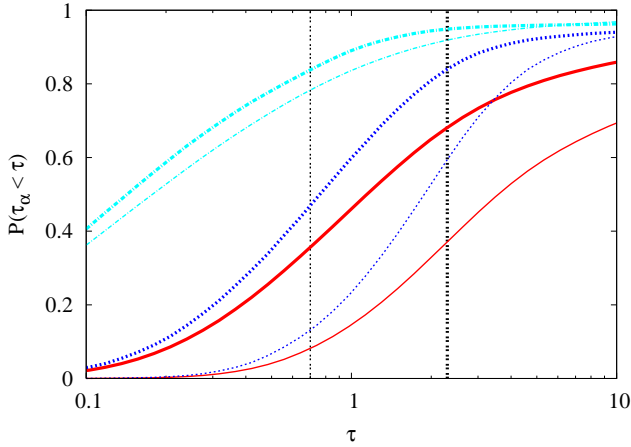


**Figure 6.** The same as Figure 5, but the curves are computed assuming model (i). *Top Panel:* Curves compare the suppression in the fiducial model in which  $L \propto m$  (source model A) to a model in which  $L$  is independent of  $m$  (source model B). *Bottom Panel:* Curves labelled “wind” assume  $v_w = 400 \text{ km s}^{-1}$  and labeled “infall” assume  $v_w = v_{\text{vir}}$ .

(i), the luminosity function in the other models is suppressed by a largely luminosity-independent factor at each  $\bar{x}_i$ . The sharp cutoff at high luminosities in this ratio owes to the finite simulation volume. Model (ii) has the largest bubbles so that the luminosity function is the least suppressed at fixed  $\bar{x}_i$ , whereas model (iii) has the smallest bubbles such that the luminosity function is the most suppressed.

We also calculated the suppression of the luminosity function for a wind model with  $v_w = 400 \text{ km s}^{-1}$ . This value for  $v_w$  is motivated by the measured average velocity offset of the Ly $\alpha$  line in strong emitters at  $z \approx 3$  (Shapley et al. 2003). This model probably yields an upper limit for the

<sup>9</sup> To perform this calculation, we use the halo field at  $z = 6.6$  in the 186 Mpc box, but we use the ionization field from the simulations of model (i), which are generally at slightly higher redshifts. This is justified because the ionization field, when comparing at fixed  $\bar{x}_i$ , is essentially independent of the redshift where this  $\bar{x}_i$  is reached (McQuinn et al. 2006).



**Figure 7.** Probability of having  $\tau_\alpha(\nu_0) < \tau$ . The thin curves are for LAEs with  $1 \times 10^{10} M_\odot < m < 2 \times 10^{10} M_\odot$ , and the thick curves are for LAEs with  $5 \times 10^{10} M_\odot < m < 1 \times 10^{11} M_\odot$ . The solid curves are for  $\bar{x}_i = 0.3$ , the dotted curves are for  $\bar{x}_i = 0.5$ , and the dot-dashed curves are for  $\bar{x}_i = 0.7$ . Assuming the observational threshold of  $m > 1 \times 10^{10} \exp(-\tau_\alpha(\nu_0)) M_\odot$ , only the emitters that contribute to the portion of the thin (thick) curves leftward of the thin (thick) vertical lines are observed.

effect of winds at  $z > 6$ , where the galaxies are much less massive than at  $z \approx 3$  and, therefore, not able to power such strong winds. A wind causes a redshift of the Ly $\alpha$  line because Ly $\alpha$  photons lose energy when they scatter off the baryons in the wind. Winds make the bubble size needed for a fixed damping wing optical depth shrink. See the curves labeled “wind” in the bottom panel in Figure 6 and compare with the other curves with the same  $\bar{x}_i$ . This comparison illustrates the effect of a winds on the luminosity function can be significant for  $\bar{x}_i \lesssim 0.3$ , but tend to be unimportant when  $\bar{x}_i \gtrsim 0.6$ . The effect of winds on the luminosity function decreases with increasing  $\bar{x}_i$  because as the bubbles grow (increasing  $\bar{x}_i$ ) the ratio of the wind velocity to the Hubble flow velocity at the bubble edge,  $v_w/(H(z_g) R_b)$ , decreases.

We also consider a model in Figure 6 where we set  $v_w = v_{\text{vir}}(m)$ , where  $v_{\text{vir}}(m)$  is the halo circular velocity at the virial radius. This model is meant to emulate analytic calculations in which resonant absorptions of infalling material can obscure all wavelengths blueward of  $\lambda(1+v_{\text{vir}}(m)/c)$  (Santos 2004; Dijkstra et al. 2007). As with winds, infall has a relatively minor effect at  $\bar{x}_i \gtrsim 0.6$ , but can reduce the suppression of the luminosity function for  $\bar{x}_i \lesssim 0.3$  (Figure 6).

The uniform suppression of the luminosity function was also found in the analytic studies of Furlanetto et al. (2006). The explanation provided in Furlanetto et al. (2006) for this uniform suppression is that the most massive (most biased) sources sit in the largest bubbles, which results in the least attenuation of their Ly $\alpha$  line. In addition, the most massive LAEs are the most luminous (at least in our model), requiring a larger  $\tau_\alpha$  than an average mass halo does to be obscured. However, at the bright end of the luminosity function, even a slight decrease in the average luminosity of the LAEs causes the luminosity function to change rapidly because of its steep slope. These effects sum to give roughly the same suppression at the bright end as at the faint end. In what follows, we investigate this explanation in detail and

quantify the effect of our assumptions on our predictions for the evolution of the luminosity function.

Figure 7 illustrates the dependence of damping wing absorption on halo mass, plotting the cumulative probability distribution of  $\tau_\alpha(\nu_0)$ ,  $P(\tau_\alpha(\nu_0) < \tau)$ . The thick curves are  $P(\tau_\alpha(\nu_0) < \tau)$  for LAEs with  $5 \times 10^{10} M_\odot < m < 1 \times 10^{11} M_\odot$ , and the thin ones are  $P(\tau_\alpha(\nu_0) < \tau)$  for LAEs with  $1 \times 10^{10} M_\odot < m < 2 \times 10^{10} M_\odot$ . This figure demonstrates that the most massive LAEs have smaller  $\tau_\alpha$  because they sit in larger bubbles than less massive LAEs. This effect decreases with increasing  $\bar{x}_i$  because the bubbles become less biased as they grow. (Compare the difference between the two solid curves –  $\bar{x}_i = 0.3$  – to the difference between the two dot-dashed curves –  $\bar{x}_i = 0.7$ .) In addition, the most massive LAEs are the most luminous in our model, which requires a larger  $\tau_\alpha$  than less massive LAEs do for their Ly $\alpha$  luminosity to fall below the survey threshold  $L_{\text{min}}$ .

Figure 7 suggests that our predictions for the luminosity function depend on the scaling of the LAE bias with luminosity. This is worrisome because this scaling is very uncertain. How dependent are our predictions on the simple mapping we assume between halo mass and Ly $\alpha$  luminosity? One plausible extreme LAE model is, rather than  $L_{\text{int,E}} \propto m$ , for  $L_{\text{int,E}}$  to be independent of halo mass. To achieve this, we kept the same halo positions but randomized the luminosities of the halos while maintaining the same luminosity function as in the fiducial Ly $\alpha$  emitter model. We refer to this source model as model B and the fiducial model as model A. The top panel in Figure 6 compares the suppression of the luminosity function in these two models. The suppression in the two models is similar, particularly for the  $\bar{x}_i = 0.62$  case. The agreement between these two models suggest that our predictions are not strongly dependent on our prescription for the luminosity function.

When  $\bar{x}_i = 0.33$ , the luminosity function is more suppressed at the high mass end for source model B than for source model A, imparting some scale dependence in the suppression for model B. The scale dependence induced by model B can be thought of as the maximum scale dependence that can be imparted by reionization since bias is uncorrelated with luminosity in this model. Therefore, it is difficult for reionization to be solely responsible for the scale-dependent suppression that may have been observed in Kashikawa et al. (2006).<sup>10</sup>

If a LAE survey suffers from significant contamination or incompleteness, these systematics will affect the normalization and shape of the luminosity function, complicating any inference that the evolution is due to neutral regions. Also, intrinsic evolution in the source properties with redshift could be difficult to distinguish from reionization. If evolution in the observed luminosity function indeed owes to reionization, then reionization will also increase the clustering of observed emitters, whereas systematic effects and intrinsic evolution cannot change the clustering in the same

<sup>10</sup> It is possible to induce more of a scale dependence in the suppression by making the luminosity function steeper than we have assumed at large luminosities. However, we find that in practice a steeper bright-end luminosity function does not result in much additional suppression. This result owes to the large dispersion in  $\tau_\alpha$  for halos of fixed  $L_{\text{obs,E}}$ .

manner. The impact of reionization on clustering is discussed in the following section.

## 5 CLUSTERING

Suppose that upcoming observations confirm the substantial decrease in  $\bar{n}_E$  between  $z = 5.7$  and  $z = 6.6$  found in Kashikawa et al. (2006). This evolution could be explained in three ways (or some combination thereof):

(i) *Decreasing Duty Cycle*: The number density of halos that host emitters is decreasing, but the observed emitters sit in halos of the same mass at  $z = 6.6$  as at  $z = 5.7$ .

(ii) *Increasing Halo Mass*: The average halo mass of LAEs is increasing between  $z = 5.7$  and  $z = 6.6$ .

(iii) *Reionization*: Patchy reionization is still occurring at  $z = 6.6$ . Neutral regions are obstructing the line of sight to some emitters, decreasing their observed abundance.

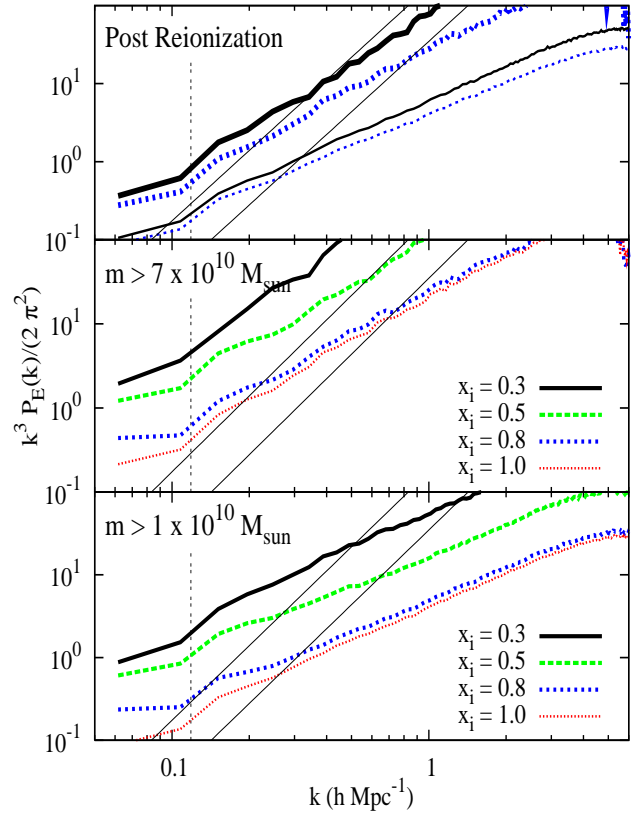
In case (i), the clustering properties of the sources will be essentially unchanged in the 150 million years between  $z = 5.7$  and  $z = 6.6$ . Only in cases (ii) and (iii) can the amount of clustering increase significantly. If a high-redshift LAE survey is able to distinguish case (ii) from case (iii), it will be capable of determining whether reionization is happening at  $z = 6.6$ .

The curves in Figure 8 are the 3-D power spectrum of  $\delta_E \equiv n_E/\bar{n}_E - 1$ , calculated from a simulation of model (i). The shot noise component of this power spectrum has been removed from these curves. Note that these curves depend only on  $m_{\min}$  and the ionization field; they do not depend on  $f_E$ . We plot the 3-D rather than the 2-D power spectrum because the 3-D power spectrum makes use of all the two-point information that is available in our simulation volume, minimizing cosmic variance. Note that all of our conclusions about the effect of reionization on the LAEs would be the same if we considered the 2-D power spectrum. The curves in the middle panel in Figure 8 represent the case that the observed LAEs have  $m \exp(-\tau_\alpha(\nu_0)) > 7 \times 10^{10} M_\odot$ . (Note that  $L_{\text{int,E}} \sim m$ .) This threshold along with  $f_E \approx 0.1$  yields the  $\bar{n}_E$  measured in the SDF at  $z = 6.6$ .

The curves in the bottom panel in Figure 8 represent the case in which LAEs are observed in less massive halos than the middle panel, halos with  $m \exp(-\tau_\alpha(\nu_0)) > 1 \times 10^{10} M_\odot$ . The curves correspond to the signal for surveys that are more sensitive than Subaru such as JWST (unless  $f_E \lesssim 0.01$ ). The LAE power spectrum at fixed  $\bar{x}_i$  has a slightly lower amplitude when computed from more abundant halos.

During reionization, the power spectrum of LAE fluctuations changes rapidly. Between  $\bar{x}_i \approx 1$  (when  $z = 6.7$  in the simulation of model (i)) and  $\bar{x}_i = 0.5$  (when  $z = 7.5$ ), the amplitude of the power spectrum increases by roughly a factor of  $\approx 3 - 4$  (see middle panel in Fig. 8). The amplitude increases by another factor of  $\approx 2$  by  $\bar{x}_i = 0.3$  (when  $z = 8$ ). If LAE surveys detect a rapid increase in the amplitude of the power spectrum with redshift, it would be difficult to attribute this to anything other than reionization.

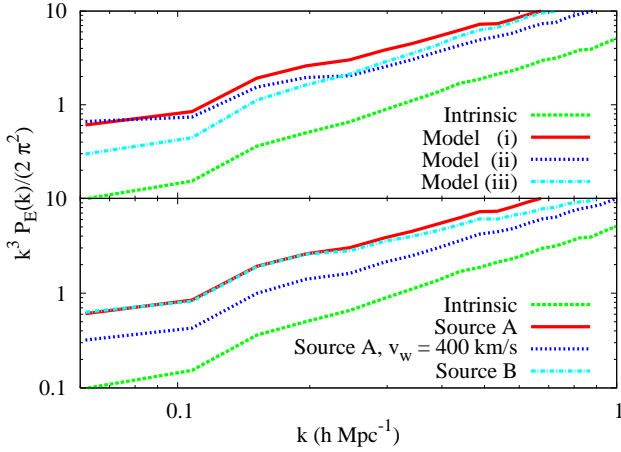
The straight, solid diagonal lines in Figure 8 represent the shot-noise power spectrum for  $\bar{n}_E = 1 \times 10^{-4} \text{ Mpc}^{-3}$  and  $\bar{n}_E = 5 \times 10^{-4} \text{ Mpc}^{-3}$ . The former  $\bar{n}_E$  is approximately the  $\bar{n}_E$  measured in the SDF photometric sample



**Figure 8.** Dimensionless 3-D power spectrum of  $\delta_E \equiv n_E/\bar{n}_E - 1$ , calculated using the 186 Mpc simulation of model (i). *Bottom Panel:* The power spectrum of emitters with  $m \exp(-\tau_\alpha(\nu_0)) > 1 \times 10^{10} M_\odot$ . All curves are from redshifts between  $z = 6.9$  and  $z = 8.3$ . *Middle Panel:* The power spectrum of emitters with  $m \exp(-\tau_\alpha(\nu_0)) > 7 \times 10^{10} M_\odot$ , and  $\bar{n}_{\text{int,E}} \approx 10^{-3} f_E \text{ Mpc}^{-3}$ . *Top Panel:* The intrinsic clustering of emitters (or, equivalently, the clustering when  $\bar{x}_i \approx 1$ ). The thick curves are if emitters can be observed in halos with  $m > 1 \times 10^{11} M_\odot$  at  $z = 6.9$  (dashed curve,  $\bar{n}_E = 1 \times 10^{-4} f_E \text{ Mpc}^{-3}$ ) and at  $z = 8.0$  (solid curve,  $\bar{n}_E = 5 \times 10^{-5} f_E \text{ Mpc}^{-3}$ ). The thin solid and dashed curves are the same but in halos with  $m > 1 \times 10^{10} M_\odot$  (with  $\bar{n}_E = 0.03 f_E \text{ Mpc}^{-3}$  at  $z = 6.9$  and  $\bar{n}_{\text{int,E}} = 0.02 f_E \text{ Mpc}^{-3}$  at  $z = 8$ , respectively). *All Panels:* The straight, vertical line corresponds to  $k = 2\pi h R^{-1}$ , where  $R = 80 \text{ Mpc}$  – roughly the angular extent of the  $z = 6.6$  LAE survey in the SDF. The diagonal solid lines are the shot-noise power spectrum for  $\bar{n}_E = 1 \times 10^{-4} \text{ Mpc}^{-3}$  and  $\bar{n}_E = 5 \times 10^{-4} \text{ Mpc}^{-3}$ . The former  $\bar{n}_E$  is approximately the value of  $\bar{n}_E$  in the SDF photometric sample.

(Kashikawa et al. 2006). Notice that for a survey with either value of  $\bar{n}_E$ , clustering can be detected on large scales provided that the survey volume is large enough (i.e., the shot-noise line is below the other curves). The vertical dashed lines in Figure 8 correspond to  $k = 2\pi h R^{-1}$ , where  $R = 80 \text{ Mpc}$  – roughly the angular extent of the  $z = 6.6$  LAE survey in the SDF. These curves imply that if reionization is happening at  $z = 6.6$ , fluctuations in the LAE field can be imaged even in current programs.

The results in the bottom two panels of Figure 8 should be contrasted with what observations would see if, instead of the increased clustering owing to reionization, it was enhanced by evolution in the intrinsic properties of the LAEs. A similar evolution in the clustering properties to the evolu-



**Figure 9.** *Top Panel:* Dimensionless 3-D power spectrum of  $\delta_E$  for three different models of reionization, calculated from snapshots that have  $\bar{x}_i = 0.5$  and assuming  $m \exp(-\tau_\alpha(\nu_0)) > 1 \times 10^{10} M_\odot$ . In addition, a curve for the intrinsic (or  $\bar{x}_i \approx 1$ ) power spectrum is included for comparison. *Bottom Panel:* Comparison of the  $\bar{x}_i \approx 0.5$  LAE power spectrum of source model A (in which  $L_{\text{int,E}} \sim m$ ) to both the LAE power spectrum of a wind model that uses source model A and  $v_w = 400 \text{ km s}^{-1}$  and to the LAE power spectrum of source model B (in which  $L_{\text{int,E}}$  is independent of  $m$ ). The curve labeled “intrinsic” uses source model A, and all curves in the bottom panel use reionization model (i).

tion caused by reionization could conceivably be produced if the bias of the sources were to increase with redshift – i.e. if the LAEs sit in more massive halos at  $z = 6.6$  than at  $z = 5.7$ .

In the top panel in Figure 8, the thin curves have a 10 times smaller  $m_{\text{min}}$  than do the thick ones. For fixed  $f_E$ , there are over 100 times more emitters in the survey from which the thin curves are calculated compared to the survey from which the thick curves are computed. A change in  $m_{\text{min}}$  that leads to a change in  $\bar{n}_E$  by a factor of over 100 yields a smaller variation in  $P_E$  than the change in  $P_E$  between when  $\bar{x}_i \approx 1$  and when  $\bar{x}_i = 0.5$ , even though  $\bar{n}_E$  differs by only a factor of 3 between these curves (fixing  $f_E$ ). Reionization causes the clustering of LAEs to evolve much more quickly than is possible with intrinsic evolution.

Let us develop a general understanding of the magnitude by which intrinsic evolution of the LAEs can change the amount of clustering. On large scales,  $P_E = b^2 P_{\text{DM}}$ , where  $b$  is the intrinsic large-scale bias and  $P_{\text{DM}}$  is the dark matter power spectrum. The large-scale bias calculated using Press-Schechter theory at  $z = 7$  is  $b^2 = 28$  for  $m_{\text{min}} = 1 \times 10^{10} M_\odot$ ,  $b^2 = 58$  for  $m_{\text{min}} = 5 \times 10^{10} M_\odot$ ,  $b^2 = 65$  for  $m_{\text{min}} = 7 \times 10^{10} M_\odot$ , and  $b^2 = 77$  for  $m_{\text{min}} = 1 \times 10^{11} M_\odot$ . The ratio of the  $P_E$  for  $\bar{x}_i \approx 1$  and for different  $m_{\text{min}}$  is typically smaller than the ratio between  $P_E$  for models in which  $\bar{x}_i$  differs by  $\Delta\bar{x}_i \approx 0.3$ .

The top panel in Figure 9 compares the power spectrum of  $\delta_E$  for reionization models (i), (ii), and (iii) and for  $\bar{x}_i = 0.5$  with  $m_{\text{min}} = 1 \times 10^{10} M_\odot$ . Model (iii) has the least power on large scales – despite having the fewest observed emitters of the three models – because it has the smallest bubbles. However, the differences between the curves for  $\bar{x}_i = 0.5$  are not as substantial as the differences between  $\bar{x}_i = 0.3$ ,  $\bar{x}_i = 0.5$ , and  $\bar{x}_i \approx 1$ . Therefore, it will be easier to constrain  $\bar{x}_i$

to  $\sim 30\%$  with emitter clustering than it will be to constrain the details of the reionization process.

The bottom panel in Figure 9 compares the power spectrum of  $\delta_E$  computed from a LAE field that uses the fiducial source model (model A) and from a LAE field that uses source model B, which was discussed in Section 4, where luminosity is independent of halo mass for all halos above  $m_{\text{min}}$ . All curves are calculated with  $m_{\text{min}} = 1 \times 10^{10} M_\odot$ . The bottom panel in Figure 9 also plots the power spectrum for a LAE model with  $v_w = 400 \text{ km s}^{-1}$  for all emitters. Winds suppress the amount of clustering.

If a LAE survey detects excess power on large scales, a skeptic might contend that this is caused by a rare large-scale structure in the survey field rather than by reionization. Fortunately, there is a simple test that may help to distinguish between these two hypotheses: Survey the field using a different selection criterion, such as H $\alpha$  emission or by the Lyman-break technique. If the reionization hypothesis is correct, less clustering is expected in the second survey, and if we cross correlate with the Ly $\alpha$  survey, again the same excess clustering will be present.<sup>11</sup>

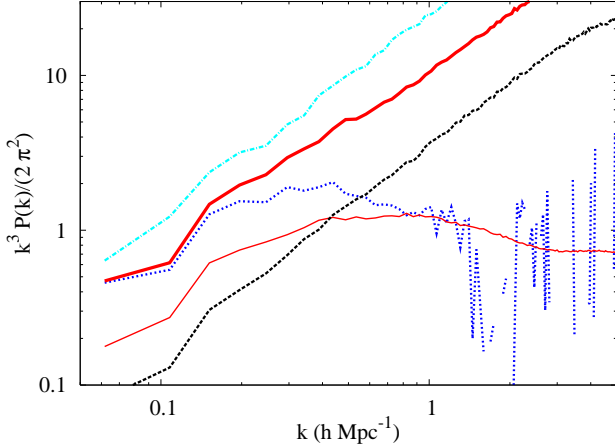
This test does not rule out the possibility that the LAEs are intrinsically more biased than the galaxies selected with the other selection criterion. This possibility may be surprising because at lower redshifts LAEs are associated with younger, less biased systems than the Lyman break galaxies. However, let us suppose that this circumstance is the case. On scales where the intrinsic bias is linear and shot noise is unimportant, the Fourier space fluctuations of the two galaxy fields should have the same phase at each  $\mathbf{k}$  if  $\bar{x}_i \approx 1$ . If this is not true, then this would be evidence for reionization. Of course, on large enough scales the emitter field,  $\delta_E$ , will have the same phase as the galaxy field,  $\delta_g$ , even during reionization, but the presence of HII bubbles will extend the range of scales over which the phases do not agree. The cross correlation coefficient  $r(k) = \langle \delta_{E,k} \delta_{g,k} \rangle / (\langle \delta_{E,k}^2 \rangle \langle \delta_{g,k}^2 \rangle)^{0.5}$  compares the phases of two fields. If  $\bar{x}_i = 0.5$  and  $m_{\text{min}} = 5 \times 10^{10} M_\odot$ , we find in model (i) that  $r \approx 0.9$  at  $k = 0.1 \text{ Mpc}^{-1}$ , whereas if  $\bar{x}_i \approx 1$  then  $r \approx 1$  at  $k = 0.1 \text{ Mpc}^{-1}$  (even if the  $m_{\text{min}}$  differ by a factor of a few between the  $\delta_E$  and  $\delta_g$  fields).

Another potential test to check whether the fluctuations in the LAE field owe in part to reionization is to note that the HII regions break the rotational isotropy of the LAE field because Ly $\alpha$  absorption depends on the ionization state of the IGM in only the line-of-sight direction.<sup>12</sup> We have

<sup>11</sup> There is a second-order effect owing to Jeans mass suppression in ionized regions inhibiting galaxies with  $m \lesssim 10^9 M_\odot$  from forming. A high redshift galaxy would be composed of many of these smaller mass galaxies owing to past merger events, and therefore its Ly $\alpha$  emission might be influenced by the local reionization history (Babich & Loeb 2006; Pritchard et al. 2007). This effect may result in reionization also influencing the clustering of all galaxies, not just LAEs. Quantifying the relevance of this effect is difficult, but it is certainly much smaller than the modulation we consider.

<sup>12</sup> Physically, a uniform distribution of LAEs and a spherical HII region will result in a non-spherical observed distribution of LAEs within the HII region, with the distribution of LAEs truncated at  $\sim 1 \text{ pMpc}$  from the edge of the HII region along the line of sight direction. In reality, the distribution of emitters and the shape of





**Figure 10.** This plot investigates the effects that shape  $k^3 P_E(k)/2\pi^2$ . The dashed curve is the intrinsic LAE power spectrum for  $m_{\min} = 1 \times 10^{10} M_\odot$ . The thick solid curve is  $k^3 P_E(k)/2\pi^2$  for  $\bar{x}_i = 0.5$  and  $m_{\min} = 1 \times 10^{10} M_\odot$ . The dotted curve is the same as the thick solid except that each of the halos has been randomly displaced in the box prior to computing the observed LAE field. The dot-dashed curve is the same as the thick solid except that the halo field is uniformly translated by some arbitrary distance relative to the ionization field (to make  $\bar{x}_i$  and  $\delta_g$  uncorrelated) prior to computing the observed LAE field. The shot noise contribution has been subtracted from the curves that represent  $P_E$ . The thin solid curve is five times the power spectrum of  $\delta_{x_i}$ , included to compare its spectral dependence with the other curves.

investigated the statistic  $\langle |\delta_E(\mathbf{k} \cdot \hat{\mathbf{n}})|^2 \rangle$  for different orientations of  $\hat{\mathbf{n}}$  and have found no significant signature of angular anisotropy in our simulation cube (neglecting redshift-space distortions owing to the peculiar velocity field).

In this section, we focused on the  $\bar{n}_E$ -weighted 3-D power spectrum. The reader might wonder why we did not use the luminosity-weighted power spectrum (or even some more general weighting) instead, which can contain more information about reionization. We find that a luminosity-weighted power spectrum results in a similar spectrum of fluctuations. The reason for this is that most emitters in a survey are near the detection threshold because of the decreasing nature of the mass function. It only takes minor attenuation for most LAEs to not be observed, such that the major source of fluctuations derives from whether an emitter is detected or not.

### 5.1 Factors that Shape the LAE Power Spectrum

Let us develop a simple model for the LAEs to understand the reason that reionization has such a large effect on  $P_E$ . We will not use this model to predict  $P_E$ , but rather to provide a framework with which to understand the different effects that contribute to  $P_E$  during reionization. For this model, we assume that only emitters in bubbles of radius  $l$  are observed, where  $l \sim 1$  pMpc. In reality,  $l$  depends on the emitter luminosity and on the minimum luminosity of the survey. We can write the observed number density of

HII regions is more complicated than in this spherical example, which reduces the magnitude of this effect.

LAEs at position  $\mathbf{x}$  as  $n_E(\mathbf{x}) = n_g(\mathbf{x}) \tilde{x}_i^\gamma(\mathbf{x})$ , where  $\tilde{x}_i$  is the ionization field smoothed over a sphere of radius  $l$ , and  $n_g$  is the local number density in galaxies that emit in Ly $\alpha$ . If a bubble is defined as a fully ionized sphere then  $\gamma = \infty$ . However, in practice regions are never fully ionized, and for our purposes it suffices to leave  $\gamma$  as a free parameter.

In this model, the correlation function of LAEs can be written as (omitting constants)

$$\langle \delta_E \delta_E' \rangle = X^2 \left( \langle \tilde{x}_i^\gamma \tilde{x}_i^{\gamma'} \rangle + 2 \langle \tilde{x}_i^\gamma \delta_g \tilde{x}_i^{\gamma'} \rangle + \langle \tilde{x}_i^\gamma \delta_g \tilde{x}_i^{\gamma'} \delta_g' \rangle \right), \quad (4)$$

where  $\delta_E$  and  $\delta_g$  are the overdensities in observed LAEs and Ly $\alpha$  emitting galaxies, respectively, and where  $X = \bar{n}_g/\bar{n}_E$ . Even though this model is simplistic, if we had the full model for this correlation function it would have a similar decomposition to the decomposition seen in the right-hand side (RHS) of equation (4).

Which terms on the RHS of equation (4) shape  $\delta_E$ ? The first term is generated only by the bubbles. On scales where this term is dominant,  $P_E$  is independent of the intrinsic clustering of the LAEs. To investigate the importance of this term, we randomly displaced all the halos in the simulation box such that the intrinsic halo field is Poissonian prior to computing the LAE field,  $\delta_E(\mathbf{x})$ . This operation makes the second and third terms in equation (4) zero for finite separations. The power spectrum of  $\delta_E$  using the displaced halo field for  $\bar{x}_i = 0.5$  is given by the dotted curve in Figure 10 (with the shot noise contribution subtracted off). Compare this curve to the thick solid curve, which represents  $P_E$  computed from the true halo field. On large scales, these curves agree fairly well, implying that the first term on the RHS of equation (4) is important at tens of Mpc scales and greater. This result qualitatively agrees with the analytic study of Furlanetto et al. (2006), which found that the bubbles dominate the large-scale emitter clustering.

The thin solid curve in Figure 10 is five times the power spectrum of  $\delta_{x_i}$  (the normalization is chosen to facilitate a comparison of the shape of this curve with the shape of other curves). As expected, the spectral shape of the thin solid and dotted curves are similar. However, unlike the thin solid curve, the dotted curve is consistent with zero at  $k \gtrsim 2 h \text{ Mpc}^{-1}$ . This difference is because sub-pMpc features in the bubbles do not affect the LAE field.

We have quantified the importance of the first term, but what about the second and third terms on the RHS of equation (4)? These terms depend in part on the covariance between  $\delta_g$  and  $x_i$ . To investigate the importance this covariance, we translated the halo field in the simulation box relative to the ionization field such that  $\delta_g$  and  $x_i$  become uncorrelated.  $P_E$  increases at all  $k$  by  $\approx 50\%$  from this operation (compare dot-dashed curve with thick solid curve in Fig. 10), and the shape of  $P_E$  is maintained.

Since  $P_E$  is not drastically changed by correlations between  $\delta_g$  and  $x_i$ , this result motivates the assumption that  $\delta_g$  and  $x_i$  are uncorrelated to understand  $P_E$ . The second term on the RHS of equation (4) is zero with this assumption and the third term becomes  $X^2 \langle \delta_g \delta_g' \rangle \langle \tilde{x}_i^\gamma \tilde{x}_i^{\gamma'} \rangle$ . With these simplifications, the Fourier transform of equation (4)

becomes

$$P_E(\mathbf{k}) = X^2 \left[ P_{\tilde{x}^\gamma \tilde{x}^\gamma}(\mathbf{k}) + \int d^3 \tilde{\mathbf{k}}' P_{gg}(\mathbf{k} - \mathbf{k}') P_{\tilde{x}^\gamma \tilde{x}^\gamma}(\mathbf{k}') \right], \quad (5)$$

where  $d^3 \tilde{\mathbf{k}} = (2\pi)^{-3} d^3 \mathbf{k}$ .

At  $k \gtrsim 1 \, h \text{Mpc}^{-1}$ ,  $P_E$  is a constant times the intrinsic power spectrum (see the curves in Fig. 8). This limit is easy to understand in this model. On these scales,  $P_{\tilde{x}^\gamma \tilde{x}^\gamma} \approx 0$  and  $P_E \approx X^2 \int d^3 \tilde{\mathbf{k}}' P_{gg}(\mathbf{k} - \mathbf{k}') P_{\tilde{x}^\gamma \tilde{x}^\gamma}(\mathbf{k}') \approx X^2 P_{gg}(k) \int d^3 \tilde{\mathbf{k}}' P_{\tilde{x}^\gamma \tilde{x}^\gamma}(\mathbf{k}')$ . The  $\tilde{x}_i^\gamma$  field is primarily composed of zeros and ones, such that  $\int d^3 \tilde{\mathbf{k}}' P_{\tilde{x}^\gamma \tilde{x}^\gamma}(\mathbf{k}') = \langle \tilde{x}^\gamma \tilde{x}^\gamma \rangle(0) \approx \langle \tilde{x}^\gamma \rangle = X^{-1}$ , where  $\langle \tilde{x}^\gamma \tilde{x}^\gamma \rangle(0)$  is the correlation function evaluated at zero separation. Therefore, equation (5) becomes  $P_E \approx X P_{gg}$ . At  $\bar{x}_i = 0.5$ , uncorrelated  $\delta_g$  and  $x_i$  results in  $X \approx 10$  in our calculations, and 10 yields the small-scale increase between  $P_{gg}$  and  $P_E$  in Figure 10 (compare dashed and dot-dashed curves).

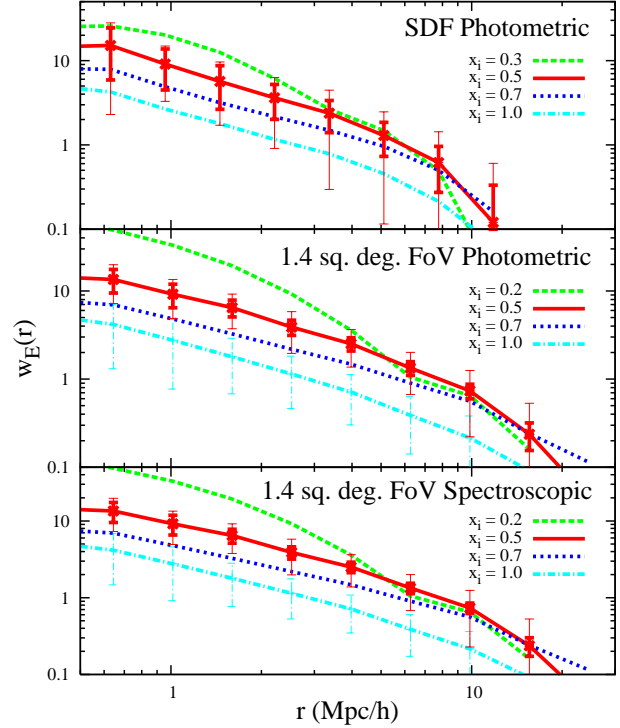
If  $P_E \propto P_{gg}$  on small scales then the proportionality constant must be  $X$  to give the correct shot noise term. In model (i), when  $\bar{x}_i = 0.5$  then  $X \approx 3$ , and when  $\bar{x}_i = 0.3$  then  $X \approx 8$ . Interestingly, these factors yield roughly the small scale power increase we see in Figure 8. Therefore, the relation  $P_E \approx X P_{gg}$  at  $k \gtrsim 1 \, \text{Mpc}^{-1}$  seems to hold in general and provides a consistency check for the reionization hypothesis, where  $X$  can potentially be estimated using a slightly lower redshift LAE survey to derive  $\bar{n}_g$ .

## 5.2 Detectability

Kashikawa et al. (2006) measured the angular correlation function  $w_E(r)$  from the SDF photometric sample of 58 emitters. Interestingly, they find that  $w_E(r)$  is consistent with no clustering (see the connected circles in Fig. 12 for this correlation function). This result may allow us to put constraints on  $\bar{x}_i$  at  $z = 6.6$ . In addition, the  $z = 6.6$  LAE sample will increase in size by a factor of five in the coming year with the Subaru/XMM-Newton Deep Survey (SXDS) (Ouchi et al. 2005).

To proceed, we estimate the mean value of  $w_E(r)$  as well as the covariance in  $w_E(r)$  between radial bins,  $C_E$ , from many mock surveys. Given the dimensions of the survey, we generate these mock catalogs in as many spatially independent volumes as our 186 Mpc simulation box allows, and we do this computation for  $\tau_\alpha(\nu_0)$  calculated along the  $\hat{i}$ ,  $\hat{j}$ , and  $\hat{k}$  directions. (For the SDF, 20 spatially independent surveys can fit into the simulation box, leading to 60 mock catalogs. In practice, we construct many times more overlapping mock catalogs than this number to obtain all the information that is available from the simulation box.) In addition, we compute  $w_E(r)$  and  $C_E$  from all the halos above  $m_{\min}$  in each mock survey region, and then we subtract out the shot noise contribution to determine the cosmological part of the covariance matrix. This procedure takes advantage of the fact that the cosmological contribution to  $w_E(r)$  and  $C_E$  does not depend on  $f_E$ , allowing us to reduce the uncertainty in our estimates for these quantities.

Figure 11 plots predictions for the correlation function of LAEs at  $z = 6.6$  and  $m_{\min} = 7 \times 10^{10} \, M_\odot$ . The pessimistic estimate in Kashikawa et al. (2006) for the contam-

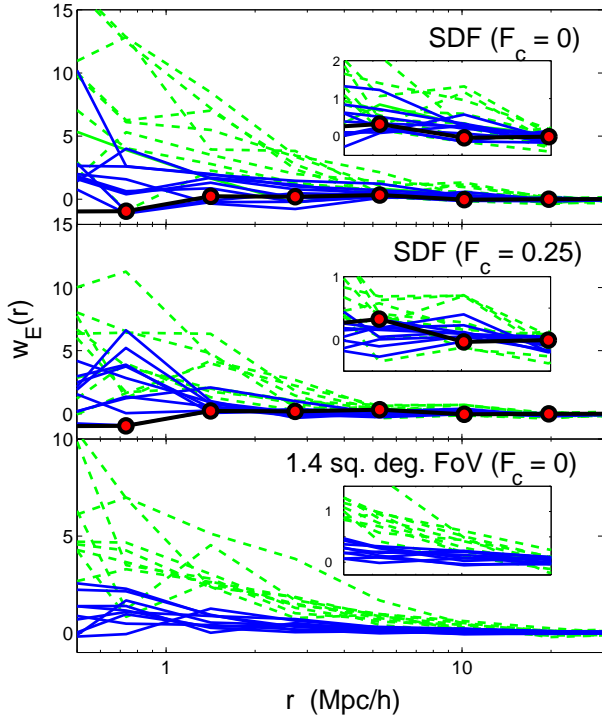


**Figure 11.** Angular correlation function of emitters at  $z = 6.6$ , assuming that observed emitters reside in halos with  $m \exp(-\tau_\alpha(\nu_0)) > 7 \times 10^{10} \, M_\odot$ . The curves in the top panel are calculated in the same volume and with the same number of emitters, 58, as the SDF photometric sample. The bottom two panels are in a volume a slightly larger volume than the upcoming 1 sq. deg. Subaru/XMM-Newton Deep Survey (SXDS), with 250 emitters in the middle panel and with 190 in the bottom one. The thick error bars owe to shot noise, and the thin owe to shot noise plus cosmic variance. To calculate these errors, we conservatively assume  $F_c = 0.25$  in the top two panels ( $F_c = 0$  in the bottom panel). Current surveys can potentially distinguish an ionized universe (the curves labeled “intrinsic”) from a universe with  $\bar{x}_i \lesssim 0.5$ .

ination fraction of the SDF photometric sample is  $F_c = 0.27$ , and Kashikawa et al. (2006) estimates that the contamination is probably closer to  $F_c = 0.16$ . We set  $F_c = 0$  for the spectroscopic survey curves (*bottom panel*) and  $F_c = 0.25$  for the photometric surveys (*middle and top panels*). We also lower the number of emitters in the mock spectroscopic surveys by  $1 - F_c$ , where  $F_c = 0.25$  is the contamination fraction in the mock photometric sample. Foreground contamination will bias the estimate for the measurement of the correlation function by the factor  $(1 - F_c)^2$ . Rather than plot biased curves for  $w_E$  in Figure 11, we instead divide the Poisson errors by the appropriate factor to account for contamination.

The top panel in Figure 11 displays the average correlation function for several clustering models, generated in mock surveys with the same dimensions as the SDF ( $34' \times 27' \times 130 \, \text{\AA}$ ) and with 58 emitters – the number of LAEs in the SDF photometric sample. The thick error bars in Figure 11 account for shot noise and the thin error bars include both shot noise and cosmic variance. Note that the cosmic variance errors in the top panel are important, par-





**Figure 12.** Angular correlation function of emitters at  $z = 6.6$  measured from independent volumes in the 186 Mpc simulation box of model (i) for  $\bar{x}_i \approx 1$  (solid curves) and for  $\bar{x}_i = 0.5$  (dashed curves). The curves in the top two panels are calculated in the same volume and with the same number of objects, 58, as the SDF photometric sample, and the curves in the bottom panel are calculated from 250 emitters in an area that is slightly larger than the upcoming SXDS. These curves assume that  $m \exp(-\tau_\alpha(\nu_0)) > 5 \times 10^{10} M_\odot$ , and there are 8 curves in each set. The thick solid line with circles in the top two panels is the 2-D correlation of the SDF photometric sample at  $z = 6.6$  (Kashikawa et al. 2006). The insets magnify the region  $4 \text{ Mpc}/h < r < 20 \text{ Mpc}/h$ .

ticularly on large scales. The impact of cosmic variance is relatively independent of the flux sensitivity of the survey. Therefore, a larger survey volume than the SDF is necessary to mitigate its effect.

However, Figure 11 suggests that current observations in the SDF can distinguish a model with  $\bar{x}_i = 0.5$  from one with  $\bar{x}_i \approx 1$ . Figure 12 illustrates more explicitly the ability to constrain reionization with current and upcoming surveys. Note that the y-axis in Figure 12 is linear rather than log as in Figure 11. The top two panels in Figure 12 show  $w_E$  measured from different locations in our box for similar survey specifications as in Figure 11 and with  $F_c = 0$  and  $F_c = 0.25$ . The dashed curves are 8 randomly selected sets of  $w_E$  for  $\bar{x}_i = 0.5$ , and the solid ones are the same for  $\bar{x}_i \approx 1$ . Both groups of curves in the top two panels in Figure 12 are measured from independent volumes in the simulation box, and each curve is calculated from 58 LAEs. The two sets of curves are fairly distinct in the top and middle panels in Figure 12, suggesting that  $\bar{x}_i < 0.5$  can be distinguished from  $\bar{x}_i \approx 1$  using current data.

The middle panel in Figure 11 shows the correlation function for mock surveys similar to the  $z = 6.6$  SXDS that

will be completed in the next year. We compute this function in a 1.4 sq. deg. area (the angular size of the 186 Mpc simulation box), with a 130 Å narrow band filter, and from 250 emitters. The SXDS is instead 1 sq. deg., but should have a similar number of emitters. Here, the uncertainty in  $w_E$  has been reduced substantially compared to the top panel. Also see the bottom panel in Figure 12. With an SXDS-like survey, even  $\bar{x}_i = 0.7$  can, on average, be distinguished from  $\bar{x}_i \approx 1$ .

The bottom panel in Figure 11 is for a luminosity-limited spectroscopic survey in the 1.4 sq. deg. area and with  $F_c = 0$ , with the same  $m_{\min}$  as the middle panel, and with the number of spectroscopically confirmed emitters reduced by the factor  $1 - 0.25$  from the mock photometric surveys in the middle panel. Since it is more expensive to perform spectroscopy, it is interesting to estimate how much a spectroscopic sample improves the potential to detect reionization. The only difference between the top panel and the middle panel is that the shot-noise errors have been reduced by the factor  $\approx (1 - F_c)$ . A spectroscopic sample is not significantly more sensitive to  $w_E$  than a photometric sample if  $F_c = 0.25$ . The benefits of a spectroscopic survey increase for larger contamination fractions. In addition, spectroscopic surveys open up a third dimension for study. For narrow band LAE surveys, the narrow third dimension does not provide much additional information, but for broad band Ly $\alpha$  surveys (as will be conducted with JWST) three dimensional clustering statistics will be the way to go.

Motivated by the appearance of large voids in the LAE field, in Appendix B we investigate the ability of void statistics to detect reionization. We find that, provided that  $F_c \approx 0$ , void statistics can distinguish models that have different  $\bar{x}_i$  with comparable significance to the correlation function.

### 5.3 Subaru Data

We have seen that a SDF-like data set can distinguish between different clustering models for the  $z = 6.6$  LAEs. Let us quantify how well the real SDF data can distinguish between these different models. We take the  $z = 6.6$  SDF photometric sample of 58 emitters presented in Taniguchi et al. (2005) and Kashikawa et al. (2006) as well as the exact survey specifications, accounting for masked regions. We then compute  $w_E(r)$  using the Landy & Szalay (1993) unbiased estimator. The connected circles in Figure 12 represent the  $z = 6.6$  correlation function of emitters presented in Kashikawa et al. (2006).

We compute the likelihoods of different theoretical models for  $w_E$  and  $C_E$  given the data (and assuming Gaussian statistics). These models are constructed in the same way as in Section 5.2. We assume a survey that has the same angular dimensions as SDF as well as a line-of-sight window function  $W(z)$  that is a Gaussian in  $\nu$  with FWHM 132 Å and that is centered at  $z = 6.6$  (such that  $m \exp(-\tau_\alpha(\nu_0)) W(z_m) > m_{\min}$  in order to be observed). This window function is a fair approximation to the window provided by the 9210 Å narrow band filter on Suprime-Cam, but we find that even if we use a tophat for the window function, our conclusions are unchanged, implying that the exact functional form of  $W(z)$  is unimportant.

Let us compare the likelihoods of models in which  $\bar{x}_i \approx$

**Table 1.**  $\chi^2 = -2 \log \mathcal{L}$  for different clustering models given the clustering in the  $z = 6.6$  SDF photometric sample. These calculations assume model (i) for reionization. The model marked “wind” assumes  $v_w = 400 \text{ km s}^{-1}$ . Note that to achieve the observed abundance at  $z = 6.6$  for  $\bar{x}_i \approx 1$  then, on average,  $f_E = 0.02$ , 0.1 and 0.2 for  $m_{\min} = 3 \times 10^{10} M_\odot$ ,  $7 \times 10^{10} M_\odot$ , and  $1 \times 10^{11} M_\odot$ , respectively.

$m_{\min} (M_\odot)$	$\bar{x}_i$	$\chi^2 (F_c = 0)$	$\chi^2 (F_c = \frac{1}{4})$
$3 \times 10^{10}$	1.0	4.9	<b>1.1</b>
	0.7	10.3	4.9
	0.5	16.6	9.9
	0.3	22.4	15.1
$7 \times 10^{10}$	1.0	10.5	4.9
	0.7	16.0	9.6
	0.5	23.6	15.9
	0.3	32.6	23.9
$1 \times 10^{11}$	1.0	14.2	7.8
	0.7	19.4	12.2
	0.5	29.3	20.7
$3 \times 10^{10}$ (wind)	0.7	8.2	3.2
	0.5	11.5	5.4
	0.3	13.8	7.5

1 to models in which the universe is significantly neutral. Table 1 summarizes our results. We find that if  $m_{\min} = 7 \times 10^{10} M_\odot$  and  $F_c = 0.25$  (adjusting  $f_E$  to yield the observed  $\bar{n}_E$ ), a universe with  $\bar{x}_i \approx 1$  is favored by the data over a universe with  $\bar{x}_i = 0.5$  at  $3.3\text{-}\sigma$  (with  $\bar{x}_i = 0.7$  at  $2.2\text{-}\sigma$ ). For a model with  $m_{\min} = 3 \times 10^{10} M_\odot$  and  $F_c = 0.25$  (which requires  $f_E = 0.02$  when  $\bar{x}_i \approx 1$ ), a universe with  $\bar{x}_i \approx 1$  is favored over a universe with  $\bar{x}_i = 0.5$  at  $3.0\text{-}\sigma$ .

Using reionization model (i), if we marginalize over  $m_{\min}$  and  $F_c$  with the priors  $m_{\min} > 3 \times 10^{10} M_\odot$  and  $F_c > 0.25$  (with  $v_w = 0$ ), the maximum likelihood model with  $\bar{x}_i \approx 1$  is favored over the maximum likelihood model with  $\bar{x}_i < 0.5$  at  $3.0\text{-}\sigma$  ( $\bar{x}_i < 0.7$  at  $1.9\text{-}\sigma$ ). If we relax the prior to  $m_{\min} > 1 \times 10^{10} M_\odot$  (such that at equality  $f_E = 4 \times 10^{-3}$ ) than  $\bar{x}_i < 0.5$  is then disfavored at  $2\text{-}\sigma$ . If we used model (ii) [model (iii)] to do the calculation discussed in the first sentence in this paragraph, a model with  $\bar{x}_i \approx 1$  is preferred over a model with  $\bar{x}_i < 0.5$  at  $2.9\text{-}\sigma$  [ $2.8\text{-}\sigma$ ]. Also, for reionization model (i) and the toy galactic wind model in which all LAEs have  $v_w = 400 \text{ km s}^{-1}$ ,  $\bar{x}_i < 0.5$  is disfavored at  $2.1\text{-}\sigma$ .

It is useful to note that the clustering data of  $z = 6.6$  LAEs can even be used to constrain the intrinsic clustering of emitters if we assume that  $\bar{x}_i \approx 1$  at  $z = 6.6$ . For example, the current data favors  $m_{\min} = 3 \times 10^{10} M_\odot$  over  $m_{\min} = 7 \times 10^{10} M_\odot$  at  $1.9\text{-}\sigma$  assuming  $F_c = 0.25$ . The current data favors  $F_c = 0.5$  over  $F_c = 0.25$  at  $1.5\text{-}\sigma$  for  $m_{\min} = 3 \times 10^{10} M_\odot$ .

In this analysis, we have assumed the statistics are Gaussian. We use 8 logarithmically-spaced radial bins that run between 0.4 and 70 Mpc to compute  $w_E$  and  $C_E$ . The bins at the smallest radii are the least Gaussian because they contain the fewest pairs of LAEs. If we discard the first two radial bins ( $r < 1.5$  Mpc), the maximum likelihood model for  $\bar{x}_i \approx 1$  can be distinguished by the SDF data from the maximum likelihood model for  $\bar{x}_i < 0.5$  at  $2.6\text{-}\sigma$  rather than

at  $3.0\text{-}\sigma$  assuming model (i) for reionization and the priors  $m_{\min} > 3 \times 10^{10} M_\odot$ ,  $F_c > 0.25$ , and  $v_w = 0$ .

We ignored the effect of survey incompleteness in the above analysis. The SDF survey is complete at  $\approx 50\%$  level above the luminosity threshold  $L_{\min}$  (and at the  $\approx 75\%$  level above  $2.5 L_{\min}$ ) (Kashikawa et al. 2006). We have investigated the importance of this effect by randomly discarding half of the objects that fall below  $1.5 L_{\min}$  and one fourth of objects between  $1.5 L_{\min} < L_{\text{obs,E}} < 2.5 L_{\min}$ , and we find that this operation changes the cosmological part of  $P_E$  by  $< 25\%$  on all scales and does not significantly alter the conclusions in this section. We find that incompleteness is degenerate with the parameter  $f_E$ .

In this analysis, we have also assumed that the shape of the luminosity function was set by the shape of the mass function. The  $z = 5.7$  and  $z = 6.6$  luminosity functions are poorly constrained at the faint end, providing good fits for faint end power-law indexes of  $\beta = 2$ , 2.5, and 3 (Shimasaku et al. 2006; Kashikawa et al. 2006). The effect of reionization on the LAEs depends on the slope of the luminosity function near  $L_{\min}$  because the emitters that fall near the detection threshold are the easiest to obscure with neutral regions. The luminosity function in our models scales as  $\beta = 2.3$  at  $m_{\min} = 1 \times 10^{10} M_\odot$ ,  $\beta = 2.6$  at  $m_{\min} = 5 \times 10^{10} M_\odot$ , and  $\beta = 2.8$  at  $m_{\min} = 1 \times 10^{11} M_\odot$ . Therefore, the range of  $m_{\min}$  that we have considered spans much of the relevant parameter space for  $\beta$ .

The calculations in this section were aimed at understanding the sensitivity to reionization of widefield surveys at  $z = 6.6$  that have already been completed or will be finished this coming year. These surveys constitute a small fraction of the total Subaru observing time in a year. However, even with these surveys, constraints can be placed on reionization. The sensitivity to reionization of a mission dedicated specifically to this science would be vastly superior. We discuss the prospects for a few upcoming LAE surveys in Section 6.

## 6 FUTURE LAE SURVEYS

We have shown that if the Universe is largely neutral at  $z \lesssim 7$ , observations like those being done on Subaru should be able to study the modulation from the HII regions. What is the prospect for upcoming surveys to detect reionization at  $z > 7$ ? To answer this question, we concentrate on three instruments: (1) the Dark Ages  $z$  Ly $\alpha$  Explorer (DAzLE), which has started observing on the Very Large Telescope (VLT), (2) the configurable multi-slit spectrograph MOSFIRE (*the Multi-Object Spectrograph For Infra-Red Exploration*) to be commissioned on the Keck telescope in 2009, and (3) NIRSpec on JWST. These represent some of the most promising instruments to target high-redshift LAEs.

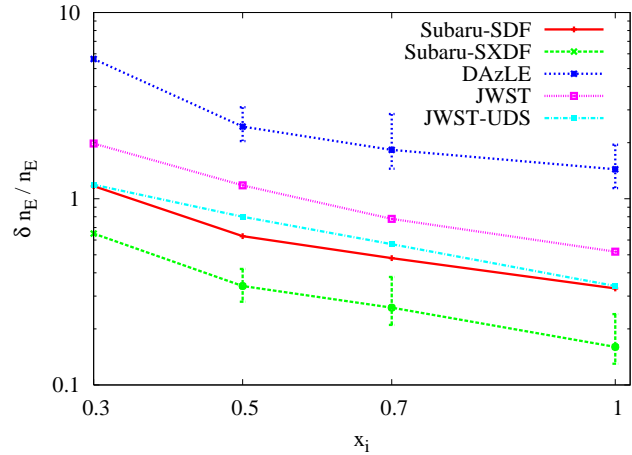
Upcoming instruments take different approaches to identify high-redshift LAEs. DAzLE uses two  $10 \text{ \AA}$ , overlapping narrow band filters in a  $6' \times 6'$  FoV. Differencing these filters makes it possible to subtract out continuum sources, leaving just the broad-line sources. The effective volume for DAzLE in a single pointing is  $1600 \text{ Mpc}^3$ , or  $\approx 1\%$  of the volume of the SDF observations. However, DAzLE’s flux sensitivity is below that of other narrow band surveys because of its extremely narrow filters. MOSFIRE uses the magnifica-

tion bias of cluster lensing caustics to improve the sensitivity to high-redshift LAEs by as much as a factor of 50. MOSFIRE is capable of simultaneously taking spectra from 45 slits, each  $7.''3$  in length. For an 80 hr observation of 8 clusters, Stark et al. (2007) estimate that these surveys could cover an effective volume of  $80 \text{ Mpc}^3$  distributed between  $z = 7.0 - 8.3$  – much smaller than the volume of narrow band samples. However, whereas DazLE will be sensitive to  $L_{\text{obs,E}} \approx 10^{42} \text{ erg s}^{-1}$  at  $z = 8$ , MOSFIRE can measure line fluxes of lensed galaxies which have  $L_{\text{obs,E}} \approx 5 \times 10^{40} \text{ erg s}^{-1}$  in a 10 hr exposure.

With the launch of JWST in 2013, the NIRSpec instrument (capable of simultaneously taking spectra of 100 objects in a  $3' \times 3'$  FoV) will be sensitive to Ly $\alpha$  luminosities that are over an order of magnitude smaller than terrestrial narrow band surveys and comparable to the effective sensitivity of a MOSFIRE caustic survey. JWST will use the imaging camera NIRCcam to select high-redshift galaxies with the Lyman break technique, and follow up with NIRSpec to select LAEs. This selection strategy will facilitate cross correlation studies to isolate the bubble-induced fluctuations (as discussed in Section 5) and allow 3-D tomography of LAEs. See Figure 3 for a mock survey of LAEs that has a JWST-like sensitivity. However, a survey with JWST that has the same area as the panels in Figure 3 would in fact require over 100 different pointings of both NIRCcam and NIRSpec to cover this 1000 sq. arc-min. field. (The box in the lower left-hand panel of Fig. 3 is the FoV for NIRSpec.) The current high-redshift program for JWST outlines a deep survey in a much smaller area, a tens of sq. arc-min. region (Gardner et al. 2006).

In general, it is difficult to make predictions for the statistical significance with which upcoming instruments can detect reionization at  $z > 7$  owing to all of the uncertainties in the source properties and their evolution. As we have seen, there is substantial uncertainty in the  $z = 6.6$  luminosity function. The extrapolation of this luminosity function to higher redshifts depends sensitively on, for example, the typical halo mass of LAEs. If  $f_E = 1$  in current LAE surveys (such that the observed LAEs are in the most massive halos), then the luminosity function will evolve more quickly with redshift than if  $f_E < 1$ . Unfortunately, the parameter  $f_E$  is not constrained to even an order of magnitude by the  $z = 6.6$  data (Dijkstra et al. 2006; Stark et al. 2007). Other factors could also shape the evolution of the Ly $\alpha$  luminosity function such as the build-up of dust or an evolving abundance of metal-poor stars within galaxies.

Despite these challenges, several studies have made estimates for  $\bar{n}_E$  as a function of redshift (Stark et al. 2007; Barton et al. 2004; Thommes & Meisenheimer 2005). For DazLE, Thommes & Meisenheimer (2005) and Stark et al. (2007) find that roughly 1 LAE should be observed at  $z = 8$  in a single 10 hr pointing of the camera. To estimate the abundance of LAEs that will be observed with MOSFIRE on Keck and NIRSpec on JWST is more difficult than for DazLE because these instruments are at least an order of magnitude more sensitive, probing a region of the luminosity function that has yet to be explored. Fortunately, an estimate for the observed number of emitters is less important for understanding the scientific impact of MOSFIRE and JWST because the limiting factor will be cosmic variance and not shot noise.



**Figure 13.** Cosmic fluctuations in the measured LAE number density as a function of  $\bar{x}_i$  for current and upcoming surveys (the variance owing to shot noise has been subtracted). Most curves assume one pointing of the relevant imaging camera/spectrograph (SDF is one pointing of Suprime-Cam), with the exceptions being those for the SXDF and UDS surveys, in which the variance is calculated for the survey volume. The DazLE and Subaru curves are computed assuming that  $m_{\text{min}} = 5 \times 10^{10} M_{\odot}$ , and those for JWST are calculated assuming that  $m_{\text{min}} = 1 \times 10^{10} M_{\odot}$ . The error bars on the DazLE and SXDF curves represent the change in  $\delta n_E / n_E$  if  $m_{\text{min}}$  is reduced or increased by a factor of 2.

Here, we make predictions for the cosmic variance in the measured LAE number density for upcoming surveys. We do not perform this calculation for a MOSFIRE cluster caustic survey owing to the complicated geometry (see Stark et al. 2007). In the context of our simple model for the LAEs, cosmic variance depends on just the parameter  $m_{\text{min}}$ . When fluctuations in the measured number density of emitters within a survey volume are on the order of unity about the average abundance  $\bar{n}_E$ , it will be difficult to understand reionization without a larger survey volume. For example, these fluctuations will prevent a precise measurement of the luminosity function.

The curves in Figure 13 represent the standard deviation in the number density of LAEs (with the shot noise component subtracted out) measured with current and upcoming surveys (and assuming no foreground contamination). Note that neutral gas increases the standard deviation significantly. The DazLE and Subaru curves are computed assuming that  $m_{\text{min}} = 5 \times 10^{10} M_{\odot}$  (which is a conservative choice and requires  $f_E < 0.1$ ), and the ones for JWST are calculated assuming  $m_{\text{min}} = 1 \times 10^{10} M_{\odot}$ . All of these curves are determined by first convolving the LAE maps with a window function that has the same volume specifications as these surveys/instruments and, then, by computing the variance of the windowed field. For JWST, which will take spectra over a broad range of wavelengths with NIRSpec, we have assumed that the line-of-sight width used to estimate  $\bar{n}_E(z)$  is 100 Mpc ( $\Delta z \sim 0.3$ ). The choice of 100 Mpc is motivated by the short duration of reionization in simulations of reionization ( $0.5 < \bar{x}_i \lesssim 1$  spans  $\Delta z \approx 1$ ) and the desire to study LAEs for a few different ionized fractions during reionization.

The curves in Figure 13 are computed from snapshots between  $z = 7$  and  $z = 8$ . The value of  $\delta \bar{n}_E / \bar{n}_E$  depends

modestly on the redshift of the survey. For example, for the SXDS, for  $m_{\min} = 5 \times 10^{10} M_{\odot}$  and  $\bar{x}_i \approx 1$  then  $\delta n_E/n_E = 0.13$  at  $z = 7$  whereas  $\delta n_E/n_E = 0.17$  at  $z = 8$ .<sup>13</sup> In addition, if we vary the value of  $m_{\min}$ , the results do not change significantly. The error bars on the DAZLE and Subaru-SXDF curves in Figure 13 represent the change in  $\delta n_E/n_E$  if the fiducial  $m_{\min}$  is reduced or increased by a factor of two.

The volume surveyed by DAZLE is much smaller than the volume of a 20 Mpc bubble. For some pointings, the surveyed volume may consist entirely of a neutral region (and zero emitters will be observed), and, for other pointings, it may fall entirely within an HII region. Figure 13 shows that in order to constrain the cosmic fluctuations in the luminosity function to 50%, DAZLE needs  $N_p \approx 10$  non-contiguous pointings when  $\bar{x}_i \approx 1$  and significantly more pointings when the Universe is largely neutral. (Note that  $(\delta n_E)_{N_p} \approx \delta n_E/\sqrt{N_p}$  for non-contiguous pointings.) Of course, this observing strategy is optimized for reducing the cosmic variance. Shot noise may be as important for DAZLE.

With JWST, one might hope to be able to place a tight constraint on the LAE luminosity function. However, for a 20% constraint, JWST needs  $N_p \approx 10$  non-contiguous pointings when  $\bar{x}_i \approx 1$  (and  $N_p \approx 50$  when  $\bar{x}_i \approx 0.5$ ).

Much of the high-redshift data for JWST will be gathered in the Ultra-Deep Survey (UDS), which will be a contiguous field spanning tens of sq. arc-min. (Gardner et al. 2006). Here, we generously assume a square field of 80 sq. arc-min ( $\approx 9$  tiles observed with both NIRCcam and NIRSpec). Figure 13 shows that this observation is far from optimal for constraining the luminosity function with  $\delta n_E/n_E \gtrsim 0.3$ . A different observing strategy is necessary for JWST to derive tight constraints on reionization with LAEs.

The largest volume LAE survey currently being conducted at  $z > 8$  is the MOIRCS Deep Survey on Subaru, which targets LAEs at  $z = 8.8$ . While not as sensitive as DAZLE, its volume of  $7' \times 4' \times 100 \text{ \AA}$  is an order of magnitude larger than that of DAZLE (Ouchi et al. 2007). If we had plotted a curve of  $\delta n_E/n_E$  for MOIRCS in Figure 13, this curve would fall between the curves for the Subaru-SDF and for DAZLE.

In this section, we have seen that cosmic variance will be a significant concern for upcoming LAE surveys at  $z > 7$ . Degree-scale surveys make the cosmic variance manageable for this science.

## 7 CONCLUSIONS

LAE surveys are probing increasingly higher redshifts. Not only will these surveys inform us about high-redshift galaxies, but they have the potential to be the first observations to unambiguously detect patchy reionization. The tens of Mpc HII regions during reionization modulate the observed distribution of LAEs and boost their observed clustering. We have shown that this effect on the angular correlation function (Section 5) or the void probability distribution (Appendix B) can be well in excess of the intrinsic clustering of

halos in the concordance cosmology. This enhanced clustering depends most strongly on  $\bar{x}_i$ , but also somewhat on the morphology of the HII regions during reionization. Observing enhanced clustering would confirm the prediction that the HII regions during reionization are large.

Even the current  $z = 6.6$  Subaru LAE survey, which has a photometric sample of only 58 LAEs in a  $0.25 \text{ deg}^2$  field, can place constraints on the reionization process. We find that the angular correlation function of the SDF photometric sample of  $z = 6.6$  LAEs favors an ionized universe over a universe with  $\bar{x}_i < 0.5$  at a  $3\text{-}\sigma$  confidence level ( $2\text{-}\sigma$  if all emitters have strong galactic winds). This constraint is both competitive with other constraints on  $\bar{x}_i$  obtained from GRBs and from LAEs (Totani et al. 2006; Malhotra & Rhoads 2006b), and it rules out the picture of the  $z = 6$  quasars expanding into a neutral IGM. In addition, this is the first constraint on the ionized fraction that consistently accounts for patchy reionization – the favored model for how the universe is ionized. Observations in the next year in the SDXS will increase the  $z = 6.6$  sample by a factor of 5 and place even stronger constraints on  $\bar{x}_i$ .

Detecting reionization through LAE clustering also offers a simple consistency check for whether the observed correlations owe to reionization: observe the same field with a second selection technique in addition to selecting objects by their Ly $\alpha$  emission. If the Ly $\alpha$ -selected galaxies show enhanced clustering relative to the galaxies selected with the other technique, the evidence for reionization is strengthened. Similarly, a comparison of the phases of Fourier modes between the two galaxy samples can be used to detect reionization.

While measuring enhanced clustering is the most fool-proof method to detect reionization with LAEs, the presence of neutral regions in the IGM also influences the properties of the Ly $\alpha$  line and the luminosity function. Observing evolution that is consistent with reionization occurring in all of these different statistics would strengthen the argument for reionization. We have made predictions for the scale of these effects. We show in Appendix A that the effect of reionization on the average line profile is less significant than most previous studies have found. This result owes to the much larger HII regions that arise from properly treating source clustering and because emitters that are not significantly obscured by the neutral regions are preferentially observed. Because of these effects, combined with the uncertainty in the astrophysical processes that determine the Ly $\alpha$  line shape, the consequences of reionization will be difficult to isolate in the line profile.

The impact of reionization on the luminosity function is a more promising diagnostic than the line profile. Its effect on the luminosity function is similar among the three reionization models we have considered, if we compare at fixed  $\bar{x}_i$ . In all models and at all  $\bar{x}_i$ , the LAE luminosity function is suppressed by a factor that is fairly constant as a function of luminosity. Our predictions for the impact of reionization on the luminosity function are fairly robust to our assumptions concerning the bias and intrinsic luminosity of emitters, but it may be difficult to distinguish evolution in the luminosity function that owes to reionization from evolution owing to changing intrinsic properties of the emitters.

It is likely that most of reionization occurs at higher redshifts than probed by current samples. Several LAE sur-

<sup>13</sup> Note that the SXDS curve in Figure 13 assumes a square survey in a  $1 \text{ sq. deg.}$  FoV. SXDS is instead a cross-shaped survey with the same volume.

veys will target  $z > 7$  LAEs in the coming years. Unfortunately, these upcoming studies are not optimal for measuring the LAE luminosity function or for detecting reionization-induced clustering, having fields of view that are too small to measure this effect precisely. Even the multi-billion dollar satellite JWST has this design flaw. JWST's enhanced sensitivity over that of current telescopes still makes it useful for this science. However, a multi-billion dollar space mission is unnecessary to understand reionization with LAEs. Widefield observations like those on Subaru, but that target  $z > 7$ , will be able to put constraints on the reionization process.

## 8 ACKNOWLEDGMENTS

We thank Nobunari Kashikawa for answering our questions about the  $z = 6.6$  SDF LAE survey, Adam Lidz for many helpful conversations, Masami Ouchi for useful discussions about the Subaru LAE surveys and helpful comments, Volker Springel for providing his Lean Gadget-2 code, Alexander Tchekhovskoy for answering MMs numerous computing questions, and Oliver Zahn for helpful comments and for providing the 2LPT displacement fields. We would also like to thank Steve Furlanetto for pointing out a mistake in a previous version of this manuscript. MM acknowledges support through an NSF graduate student fellowship. The authors are also supported by the David and Lucile Packard Foundation, the Alfred P. Sloan Foundation, and grants AST-0506556 and NNG05GJ40G.

## REFERENCES

- Babich D., Loeb A., 2006, ApJ, 640, 1  
 Barton E. J., et al., 2004, ApJ, 604, L1  
 Becker G. D., Rauch M., Sargent W. L. W., 2006, astro-ph/0607633  
 Becker R. H., et al., 2001, AJ, 122, 2850  
 Bolton J. S., Haehnelt M. G., 2007, MNRAS, 374, 493  
 Casali M., et al., 2006, in Ground-based and Airborne Instrumentation for Astronomy. Edited by McLean, Ian S.; Iye, Masanori. Proceedings of the SPIE, Volume 6269, pp. (2006). HAWK-I: the new wide-field IR imager for the VLT  
 Croton D. J., et al., 2004, MNRAS, 352, 828  
 Cuby J.-G., Hibon P., Lidman C., Le Fèvre O., Gilmozzi R., Moorwood A., van der Werf P., 2007, A&A, 461, 911  
 de Vos M., 2004, <http://www.lofar.org/PDF/LOFAR-P1-Baseline2004.pdf>  
 Dekel A., Woo J., 2003, MNRAS, 344, 1131  
 Dijkstra M., Haiman Z., Spaans M., 2006, ApJ, 649, 14  
 Dijkstra M., Lidz A., Wyithe S., 2007, Submitted to MNRAS (astro-ph/0701667)  
 Fan X., et al., 2006, AJ, 132, 117  
 Furlanetto S. R., McQuinn M., Hernquist L., 2005, MNRAS, pp 1043+  
 Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Physics Reports, 433, 181  
 Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004a, ApJ, 613, 1  
 Furlanetto S. R., Hernquist L., Zaldarriaga M., 2004c, MNRAS, 354, 695  
 Furlanetto S. R., Sokasian A., Hernquist L., 2004b, MNRAS, 347, 187  
 Furlanetto S. R., Zaldarriaga M., Hernquist L., 2006, MNRAS, 365, 1012  
 Gardner J. P., et al., 2006, Space Science Reviews, 123, 485  
 Haiman Z., 2002, ApJ, 576, L1  
 Haiman Z., Cen R., 2005, ApJ, 623, 627  
 Hansen M., Oh S. P., 2006, MNRAS, 367, 979  
 Horton A., Parry I., Bland-Hawthorn J., Cianci S., King D., McMahon R., Medlen S., 2004, PROC.SPIE INT.SOC.OPT.ENG., 5492, 1022  
 Hui L., Haiman Z., Zaldarriaga M., Alexander T., 2002, ApJ, 564, 525  
 Iliev I. T., Shapiro P. R., Raga A. C., 2005, MNRAS, 361, 405  
 Iye M., et al., 2007, Nature, 443, 186  
 Kashikawa N., et al., 2006, ApJ, 648, 7  
 Kauffmann G., et al., 2003, MNRAS, 341, 33  
 Keating B., Miller N., 2006, New Astron. Rev., 50, 184  
 Kodaira K., et al., 2003, PASJ, 55, L17  
 Landy S. D., Szalay A. S., 1993, ApJ, 412, 64  
 Lidz A., McQuinn M., Zaldarriaga M., Hernquist L., Dutta S., 2007, astro-ph/0703667  
 Lidz A., Oh S. P., Furlanetto S. R., 2006, ApJ, 639, L47  
 Liu J., Bi H., Feng L.-L., Fang L.-Z., 2006, ApJ, 645, L1  
 Loeb A., Rybicki G. B., 1999, ApJ, 524, 527  
 Madau P., Rees M. J., 2000, ApJL, 542, L69  
 Malhotra S., Rhoads J. E., 2004, ApJL  
 Malhotra S., Rhoads J. E., 2006a, ApJL, 647, L95  
 Malhotra S., Rhoads J. E., 2006b, ApJL, 647, L95  
 McPherson A. M., et al., 2006, in Ground-based and Airborne Telescopes. Edited by Stepp, Larry M.. Proceedings of the SPIE, Volume 6267, pp. (2006). VISTA: project status  
 McQuinn M., Furlanetto S. R., Hernquist L., Zahn O., Zaldarriaga M., 2005, ApJ, 630, 643  
 McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2006, Submitted to MNRAS (astro-ph/0610094)  
 McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815  
 Mesinger, A., & Haiman, Z. 2004, ApJL, 611, L69  
 Mesinger, A., & Haiman, Z. 2007, ApJ, 660, 923  
 Miralda-Escude J., 1998, ApJ, 501, 15  
 Morales M. F., Hewitt J., 2004, ApJ, 615, 7  
 Neufeld, D. A. 1991, ApJL, 370, L85  
 Osterbrock D. E., 1989, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. Sausalito, CA: University Science Books  
 Ouchi M., et al., 2005, ApJ, 620, L1  
 Ouchi M., Tokoku C., Shimasaku, K., Ichikawa, T. 2007, ASP Conf. Ser. in press  
 Page L., et al., 2006, astro-ph/0603450  
 Partridge R. B., Peebles P. J. E., 1967, ApJ, 147, 868  
 Pen U. L., Wu X. P., Peterson J., 2005, Submitted to CJAA (astro-ph/0404083)  
 Pritchard, J. R., Furlanetto, S. R., & Kamionkowski, M. 2007, MNRAS, 374, 159  
 Rhoads J. E., et al., 2004, ApJ, 611, 59  
 Santos M. G., et al., 2003, ApJ, 598, 756  
 Santos M. R., 2004, MNRAS, 349, 1137

- Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, *ApJ*, 588, 65
- Sheth, R. K., & Lemson, G. 1999, *MNRAS*, 305, 946
- Sheth R. K., Tormen G., 2002, *MNRAS*, 329, 61
- Shimasaku, K., et al. 2006, *PASJ*, 58, 313
- Sokasian A., Abel T., Hernquist L., Springel V., 2003, *MNRAS*, 344, 607
- Sokasian A., Abel T., Hernquist L. E., 2001, *New Astronomy*, 6, 359
- Sokasian A., Yoshida N., Abel T., Hernquist L., Springel V., 2004, *MNRAS*, 350, 47
- Spergel D. N., et al., 2003, *ApJS*, 148, 175
- Spergel D. N., et al., 2006, *astro-ph/0603449*
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 312
- Stark D. P., Ellis R. S., Richard J., Kneib J.-P., Smith G. P., Santos M. R., 2007, *astro-ph/0701279*
- Stark D. P., Loeb A., Ellis R. S., 2007, *astro-ph/0701882*
- Taniguchi Y., et al., 2005, *PASJ*, 57, 165
- Tapken C., Appenzeller I., Noll S., Richling S., Heidt J., Meinkoehn E., & Mehlert D. 2007, *astro-ph/0702414*
- Tasitsiomi, A. 2006, *ApJ*, 645, 792
- Thommes E., Meisenheimer K., 2005, *A&A*, 430, 877
- Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, *PASJ*, 58, 485
- White R. L., Becker R. H., Fan X., Strauss M. A., 2003, *AJ*, 126, 1
- White S. D. M., 1979, *MNRAS*, 186, 145
- Willis J. P., Courbin F., 2005, *MNRAS*, 357, 1348
- Wyithe J. S. B., Loeb A., 2006, *ApJ*, 646, 696
- Wyithe S., Loeb A., Carilli C., 2005, *ApJ*, 628, 575
- Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, *ApJ*, 654, 12
- Zahn O., Zaldarriaga M., Hernquist L., McQuinn M., 2005, *ApJ*, 630, 657
- Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, *ApJ*, 608, 622

## APPENDIX A: LINE PROPERTIES

Several studies have attempted to understand the impact that reionization has on the profile of Ly $\alpha$  lines using simple analytic models (Miralda-Escude 1998; Santos 2004; Haiman & Cen 2005; Furlanetto et al. 2004c; Dijkstra et al. 2007). In this section, we investigate the effect of reionization on the line profile using simulations. We show that even when a significant fraction of the universe is neutral, damping wing absorption owing to neutral patches in the IGM tends to have a small effect on the average line shape. We use Method 1 described in Section 2.2 to calculate the intrinsic line profile.

For simplicity, Method 1 sets the intrinsic width of the line to be  $\Delta\nu = \nu_\alpha v_{\text{vir}}/c$ , where  $v_{\text{vir}}$  is the circular velocity of the host halo at the virial radius. This  $\Delta\nu$  is the width one expects if the typical column density of HI inside the galaxy,  $N_{\text{HI}}$ , is low enough such that ionizing photons can escape from the galaxy ( $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ ) and, for the massive galaxies of interest, even if the column density is as high as  $N_{\text{HI}} = 10^{19-20} \text{ cm}^{-2}$  (Dijkstra et al. 2007). This  $\Delta\nu$  is also consistent with the emergent line profile of a 2-phase ISM in which dense clouds of HI (with dust inside) scatter

Ly $\alpha$  photons. A 2-phase structure may be necessary for the transmission of Ly $\alpha$  photons in the presence of dust (Neufeld 1991; Hansen & Oh 2006). We do not attempt to parameterize all of the complicated processes that determine the width and shape of the intrinsic line, such as the geometry of the galaxy and its bulk motions. Such processes can add additional features to the line profile and make it even more difficult to disentangle the effect of reionization.

In our calculations, the observed line profiles differ from one another primarily because of two effects: (1) The extent along the line of sight from the emitter to the HII bubble edge. The extent of the HII region determines the strength of the damping wing absorption. This effect is the most important for our study. (2) The amount of infall of gas around the emitter. This dictates how much of the red side of the line is absorbed as some photons on the red side must redshift through resonance because of the Doppler effect, typically resulting in absorption.<sup>14</sup> (Outflows – which we incorporate with a simple prescription – have the opposite effect, allowing more of the blue side of the line to be transmitted.) The effect of infall is only crudely modeled here because these computations are performed on a 0.36 Mpc grid, whereas the virial radius of a  $5 \times 10^{10} M_\odot$  halo is 0.09 Mpc at  $z = 7$ . This results in an under-prediction of the effect of infall.<sup>15</sup>

The quality of the spectrum from a single  $z > 6$  LAE for current ground-based telescopes is never good enough to constrain in detail the shape of a single Ly $\alpha$  line. Previous LAE surveys have averaged the line profile from all of their emitters to generate a higher quality effective line profile. Figure A1 shows the average line profiles calculated from snapshots with  $\bar{x}_i = 0.3$  (*top panel*),  $\bar{x}_i = 0.5$  (*middle panel*), and  $\bar{x}_i = 0.7$  (*bottom panel*) for  $z = 8.2$ ,  $z = 7.7$  and 7.3, respectively. We have normalized the integral over these curves to unity. The thick dashed curves are the average profile for LAEs with  $0.5 L_{\text{int,E}} (2 \times 10^{10} M_\odot) < L_{\text{obs,E}} < 0.5 L_{\text{int,E}} (4 \times 10^{10} M_\odot)$  for the quoted  $\bar{x}_i$ , whereas the thick solid curves are for the same luminosity range but with  $\bar{x}_i = 1$ .<sup>16</sup> Because of the exponential falloff in the halo mass function at large  $m$ , if we eliminated the upper bound on  $L_{\text{obs,E}}$  and included more luminous LAEs in the average, the line profiles would not be significantly affected.

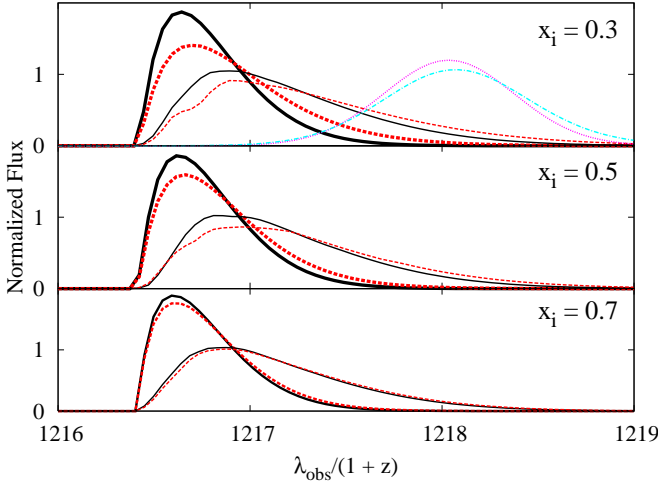
Note that the differences between the dashed and solid curves in Figure A1 are small, particularly for the  $\bar{x}_i = 0.5$  and  $\bar{x}_i = 0.7$  cases. The reason why reionization does not have more of an effect on the line profiles is because the average bubble size in which *observed* LAEs lie is larger than

<sup>14</sup> The blue side of the line is almost always entirely absorbed in our calculations. While most of the blue side of the line will typically be absorbed in reality, our calculations overestimate this effect because the simulations do not resolve the scale on which pressure smooths out density fluctuations. This lack of resolution artificially suppresses the transmission. The resolution is sufficient to capture the most important effect for this paper, the damping wing absorption.

<sup>15</sup> Analytic studies typically assume that the infall is maximal at the virial radius and equal to the circular velocity of the halo (Santos 2004; Dijkstra et al. 2007). Galactic winds can collisionally ionize the hydrogen within up to  $\sim 0.2$  pMpc of a galaxy, eliminating the importance of nearby infall (Santos 2004).

<sup>16</sup> The 0.5 that appears in these bounds is there to approximate the effect of resonant absorption, which typically absorbs all frequencies blue-ward of  $\nu_0$ .



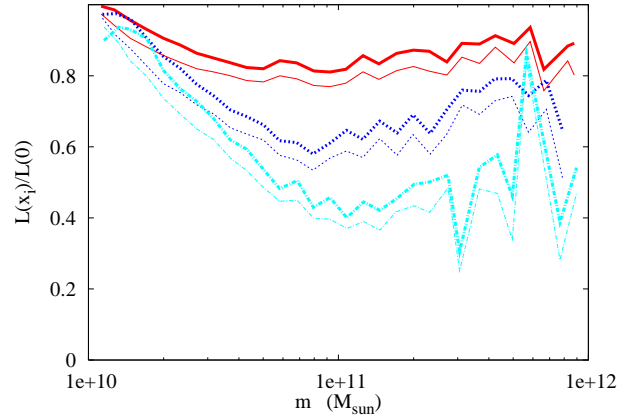


**Figure A1.** Observed line profile computed from simulation outputs in which  $\bar{x}_i = 0.3$  (*top panel*),  $\bar{x}_i = 0.5$  (*middle panel*), and  $\bar{x}_i = 0.7$  (*bottom panel*) for  $z = 8.2$ ,  $z = 7.7$ , and  $z = 7.3$ , respectively. All curves have been normalized such that their integral is unity to facilitate comparison of the line profile (see Fig. A2 for the relative normalization). The thick dashed and thin solid curves are the average profile of LAEs with  $0.5 L_{\text{int,E}}(2 \times 10^{10} M_{\odot}) < L_{\text{obs,E}} < 0.5 L_{\text{int,E}}(4 \times 10^{10} M_{\odot})$  for the quoted  $\bar{x}_i$  whereas the thick solid curves are the same for the post-reionization, ionized-IGM case. The thin dashed and thin solid curves are the same as the thick curves but for the LAEs with  $0.5 L_{\text{int,E}}(1 \times 10^{11} M_{\odot}) < L_{\text{obs,E}} < 0.5 L_{\text{int,E}}(2 \times 10^{11} M_{\odot})$ . In the top panel, the dotted and dot-dashed curves that are offset to the right are respectively the  $\bar{x}_i \approx 1$  and the  $\bar{x}_i = 0.3$  average spectrum for a toy wind model discussed in the text. These calculations are performed using the 94 Mpc simulation of Model (i).

a pMpc for the  $\bar{x}_i$  shown in Figure A1. Previous calculations assumed that each LAE was responsible for ionizing its own bubble, resulting in  $R_b < 1$  pMpc (eqn. 2) and, therefore, a larger effect of the damping wing absorption on the observed line profile (e.g., Santos 2004). One might wonder why reionization has any affect on the LAE luminosity function and correlation function (as seen in the body of this paper) since the average line profile is not strongly affected. Even though the line profile is not changed, the average transmission is decreased by reionization, and many emitters are not observed because of reduced transmission.

Galactic winds/outflows, which are found to be very prevalent in lower redshift LAEs, allow more of the blue side of the line to be transmitted. Here we adopt a simple toy model for galactic winds which is meant to exaggerate their potential effect. We make the same assumptions for the intrinsic line profile as before, but we redshift the entire line by the velocity of the wind, which is taken to be  $400 \text{ km s}^{-1}$ . This value for  $v_w$  is motivated by the average velocity offset of the Ly $\alpha$  line in strong emitters at  $z \approx 3$  (Shapley et al. 2003). Since galaxies are less massive at  $z > 6$  than at  $z \approx 3$ , the galaxies probably cannot drive such powerful winds and  $v_w = 400 \text{ km s}^{-1}$  can be thought of as an upper bound on  $v_w$ . Furthermore, it could be the case that only a fraction of the Ly $\alpha$  photons are scattered by the wind. The difference between dotted and dot-dashed curves in Figure A1 illustrate the effect of reionization in this wind model. Winds reduce the importance of damping wing absorption.

Current narrow band surveys most likely have spectro-



**Figure A2.** Ratio of the average observed luminosity during reionization to the average observed luminosity for an ionized universe as a function of halo mass. The dot-dashed, dotted, and solid thick curves are computed from simulation snapshots of Model (i) which have  $\bar{x}_i = 0.3$  ( $z = 8.2$ ),  $\bar{x}_i = 0.5$  ( $z = 7.7$ ), and  $\bar{x}_i = 0.7$  ( $z = 7.3$ ), respectively, and assuming a survey that is sensitive to  $L_{\text{obs,E}} > 0.5 L_{\text{int,E}}(1 \times 10^{10} M_{\odot})$ . These curves are computed with Method 1 in Section 2. The thin curves are the average of  $\exp(-\tau_{\alpha}(\nu_0))$  (as is computed for Method 2 in Section 2), using the emitters that are observed in the calculation with Method 1. The similarity between the thin and thick curves justifies the use of Method 2.

scopically confirmed LAEs that have  $m \approx 10^{11} M_{\odot}$  rather than  $m \approx 2 \times 10^{10} M_{\odot}$ , as considered above. (Only the brightest emitters in a narrow band survey can be spectroscopically confirmed.) If we take instead emitters within the range  $0.5 L_{\text{int,E}}(1 \times 10^{11} M_{\odot}) < L_{\text{int,E}} < 0.5 L_{\text{int,E}}(2 \times 10^{11} M_{\odot})$  (*thin curves in Fig. A1*), the differences between the line profiles for  $\bar{x}_i \approx 1$  and  $\bar{x}_i \approx 0.5$  are even smaller than in the case previously examined, owing to these more biased sources sitting preferentially in the largest HII regions.

In our computation, the FWHM of the observed line profiles for LAEs that have  $m \approx 2 \times 10^{10} M_{\odot}$  is  $\approx 4 \text{ \AA}$  (*thick curves in Fig. A1*) and for those that have  $m \approx 10^{11} M_{\odot}$  is  $\approx 8 \text{ \AA}$  (*thin curves*). The FWHM of the average line of the spectroscopically confirmed  $z = 6.6$  emitters is  $\approx 10 \text{ \AA}$ , slightly larger than the FWHM in our calculations (Kashikawa et al. 2006). Interestingly, Kashikawa et al. (2006) finds a weak anti-correlation between the FWHM of the line and  $L_{\text{obs,E}}$ . In our simple model for the LAEs, in which  $\Delta\nu \sim m^{1/3}$ , there should be a correlation. Haiman & Cen (2005) suggest that an anti-correlation may be a signature of reionization. However, to reach this result, Haiman & Cen (2005) assume that the intrinsic widths of the Ly $\alpha$  lines are the same for emitters of all  $L_{\text{int,E}}$  and that the HII region around an emitter is created just by this emitter. The anti-correlation result from Haiman & Cen (2005) depends on these dubious assumptions. There is probably too much uncertain astrophysics to understand reionization through the correlation of the FWHM with  $L_{\text{obs,E}}$ . The weak anti-correlation between FWHM with  $L_{\text{obs,E}}$  that is observed (with low statistical significance) may indicate that there is more dispersion in the mapping between  $m$  and  $L_{\text{obs,E}}$  than we have assumed in this work or that winds are important in some  $z = 6.6$  emitters. As discussed in previous sections, increasing the

dispersion in the mapping between  $m$  and  $L_{\text{obs,E}}$  does not significantly affect our conclusions pertaining to the LAE luminosity function and to LAE clustering.

The emitters in a survey whose line profiles are most strongly affected by reionization may constitute a small subsample of the LAEs in a survey. Therefore, it might be more fruitful to look at individual lines rather than the average line profile to detect reionization. However, even if a survey has sufficient signal to noise to study the Ly $\alpha$  line from a single emitter, it will be challenging to isolate the impact of damping wing absorption from other effects.

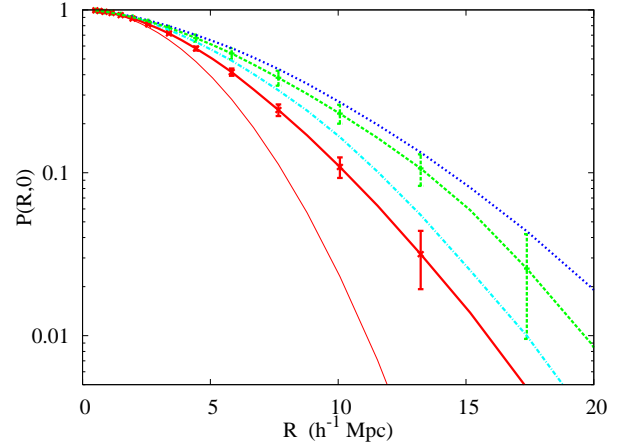
Thus far, we have considered the effect of a single model for reionization, model (i), on the line profile. In model (iii), since the bubbles are smaller, the average line profile at fixed  $\bar{x}_i$  is slightly more altered by reionization than in model (i). The opposite is true in model (ii), in which the bubbles are larger. However, the differences in the line profiles between the models at fixed  $\bar{x}_i$  are comparable to or smaller than the differences between the line profiles at  $\bar{x}_i = 0.3, 0.5$ , and  $0.7$  in model (i).

Figure A1 ignored the effect of reionization on the transmission of the Ly $\alpha$  line and just illustrated its effect on the line profile. The thick curves in Figure A2 represent the ratio of the average of  $L_{\text{obs,E}}$  at  $\bar{x}_i$  to the same but at  $\bar{x}_i \approx 1$ . Only emitters with  $L_{\text{obs,E}} > 0.5 L_{\text{int,E}} (1 \times 10^{10} M_\odot)$  are included in these averages. This ratio yields approximately the factor  $\exp(-\tau_\alpha(\nu_0))$  – the strength of damping wing absorption. The thick dot-dashed curve is for  $\bar{x}_i = 0.3$ , the thick dotted curve is for  $\bar{x}_i = 0.5$ , and the thick solid curve is for  $\bar{x}_i = 0.7$ . As  $\bar{x}_i$  decreases, the average  $L_{\text{obs,E}}$  also decreases owing to smaller bubbles and increased damping wing absorption. At low  $m$ , the decrease in  $L_{\text{obs,E}}$  with  $\bar{x}_i$  is smaller than at large  $m$ . This result might appear surprising since more massive halos sit in larger bubbles, which would result in the opposite tendency. However, this trend is simply a selection effect owing to the luminosity threshold of this mock survey. In addition, this selection effect illustrates why the line profiles are not affected as substantially as one might expect: Owing to the nature of the luminosity function, most emitters sit near the luminosity threshold. In order for these emitters to be observed, they must sit in very large HII regions such that the effect of damping wing absorption is small.

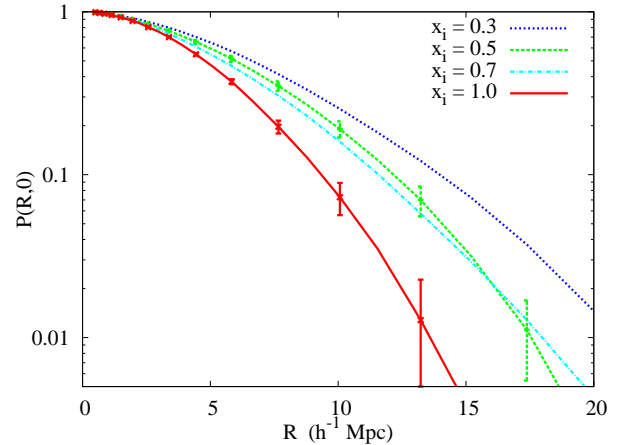
## APPENDIX B: VOID STATISTICS

The power spectrum may not be the optimal statistic to pick out the modulation of LAEs owing to the bubbles. One distinctive feature of the LAE field during reionization is that there are large voids owing to the large neutral regions (see Figs. 2 and 3). This feature motivates using the void probability distribution function  $P(R, 0)$  to measure the effect of reionization, as was done in Kashikawa et al. (2006). Void statistics have a long history of being used to interpret galaxy surveys (e.g., White 1979 and Croton et al. 2004).

$P(R, 0)$  is the probability that a circle of radius  $R$  around a point in the survey field does not contain any galaxies. For a Poisson distribution of galaxies,  $P(R, 0) = \exp(-\pi R^2 \bar{\Sigma}_E)$ , where  $\bar{\Sigma}_E$  is the surface density of LAEs. A generalization of this statistic is the probability that a region



**Figure B1.** 2-D void probability distribution function  $P(R, 0)$  at  $z = 7.5$  computed from a simulation of model (i). Each curve is calculated by averaging five independent surveys of 200 emitters in a volume of  $186 \text{ Mpc} \times 186 \text{ Mpc} \times 130 \text{ \AA}$ . The error bars are the standard deviation on the curves, calculated from five mock surveys in independent volumes within the simulation box. The thin solid curve is  $P(R, 0)$  for a purely Poisson distribution. The thick solid and dot-dashed curves are  $P(R, 0)$  for  $\bar{x}_i \approx 1$  with LAEs in halos with  $m > 5 \times 10^{10} M_\odot$  ( $f_E = 0.15$ ) and with  $m > 1 \times 10^{11} M_\odot$  ( $f_E = 0.6$ ), respectively. The long dashed [short dashed] curves are the ionized case for  $\bar{x}_i = 0.5$  and  $m > 3 \times 10^{10} M_\odot$  ( $f_E = 0.15$ ) [for  $\bar{x}_i = 0.5$  and  $m > 5 \times 10^{10} M_\odot$  ( $f_E = 0.5$ )].



**Figure B2.** The same as B1 but the curves are calculated at  $z = 6.6$  and  $m_{\text{min}} = 5 \times 10^{10} M_\odot$ . This figure compares the  $P(R, 0)$  in reionization model (i) for different  $\bar{x}_i$ .

contains  $N$  galaxies,  $P(R, N)$ , and, for a Poisson distribution,  $P(R, N) = (\pi R^2 \bar{\Sigma}_E)^N \exp(-\pi R^2 \bar{\Sigma}_E) / N!$ .

Figure B1 displays  $P(R, 0)$  computed from the  $186 \text{ Mpc}$  simulation of model (i). Each curve is calculated by averaging five independent surveys of 200 emitters in a volume of  $186 \text{ Mpc} \times 186 \text{ Mpc} \times 130 \text{ \AA}$  (a little larger than the volume of the SDSS), yielding  $\bar{n}_E = 1.5 \times 10^{-4} \text{ Mpc}^{-3}$ . The thin solid curve is  $P(R, 0)$  for a Poisson distribution. The thick solid and dot-dashed curves are  $P(R, 0)$  of an ionized universe with LAEs in halos with  $m > 5 \times 10^{10} M_\odot$  and with  $m > 1 \times 10^{11} M_\odot$ , respectively. The dashed (dotted curves) are the ionized case with  $\bar{x}_i = 0.5$  and with  $m > 3 \times 10^{10} M_\odot$ .

( $m > 5 \times 10^{10} M_{\odot}$ ). We adjust the duty cycle to fix the number of emitters, requiring  $f_E = 0.5$  for the dotted curve,  $f_E = 0.15$  for the dashed curve,  $f_E = 0.6$  for the dot-dashed, and  $f_E = 0.15$  for the thick solid.

The error bars in Figure B1 are the standard deviation from a sample of five independent mock surveys with the specifications given above. Note that the error bars in different bins are correlated. From the errors, we see that we can distinguish the models in this plot at  $> 1\text{-}\sigma$  confidence level, and, in particular, we can distinguish the case with reionization from those without reionization. This is promising because the dot-dashed curve assumes that the LAEs are in the rarest, most clustered halos, and this curve is below those with  $\bar{x}_i = 0.5$ . Clustering of LAEs owing to reionization generates large-scale voids to a much larger degree than the intrinsic clustering of halos in  $\Lambda$ CDM.

Figure B2 shows  $P(R, 0)$  for different  $\bar{x}_i$  calculated from reionization model (i). These curves are calculated assuming 200 LAE,  $m_{\min} = 5 \times 10^{10} M_{\odot}$ , and for a survey comparable in volume to the SXDS. As with the correlation function,  $P(R, 0)$  can distinguish the four different  $\bar{x}_i$  in this figure. Qualitatively, the significance with which  $P(R, 0)$  enables one to distinguish between these different  $\bar{x}_i$  appears to be comparable to the significance the correlation function affords.

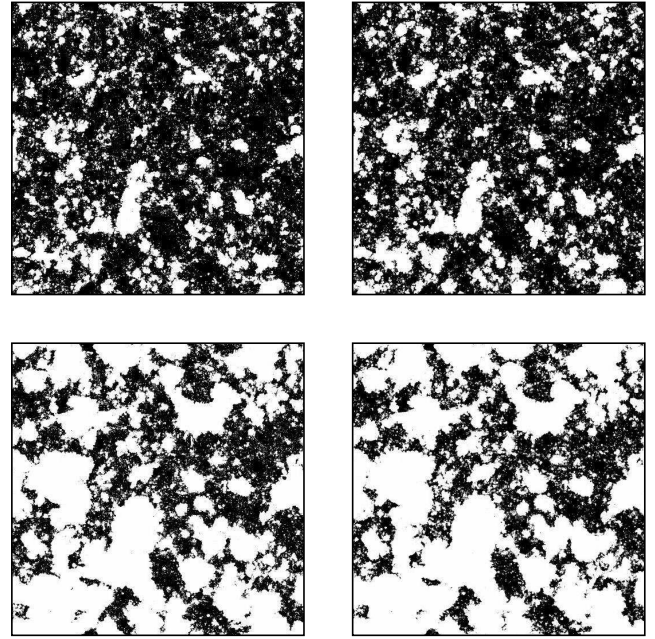
Thus far, we have assumed no contamination from foreground galaxies, which is certainly not the case for photometric LAE surveys. Unfortunately, a luminosity-limited, widefield spectroscopic LAE survey at  $z > 6$  has not been conducted. If there is significant contamination in the survey, then the measured value of  $P(R, 0)$  will be significantly biased, suppressed (on average) by the factor  $\exp(-\pi R^2 \bar{\Sigma}_{\text{cont}})$  where  $\bar{\Sigma}_{\text{cont}}$  is the density of contaminating galaxies (assuming that the contaminants are uncorrelated). If  $\bar{\Sigma}_{\text{cont}}/\bar{\Sigma}$  is appreciable, this suppression is significant, especially at large  $R$ . Kashikawa et al. (2006) measured  $P(R, 0)$  from the SDF photometric sample of LAE at  $z = 6.6$ . They found that the measured  $P(R, 0)$  is consistent with a random sample. This conclusion may owe to contamination biasing their measurement.

We can attempt to alleviate this issue of foreground galaxy contamination biasing  $P(R, 0)$  by instead computing the statistics  $P(R, 1)$ ,  $P(R, 2)$ , etc., which are progressively less biased by this effect. In this vein, we can do the exact opposite and use peaks rather than voids to probe reionization. As long as these peaks are sufficiently rare, they will exist only in a distribution of emitters that is clustered. In addition, peaks in the emitter distribution are a distinctive feature of Ly $\alpha$  maps during reionization (see Figs. 2 and 3).

We have investigated the statistic

$$P_p(R, \gamma) = \sum_{n=\text{floor}(\gamma \pi R^2 \bar{\Sigma})}^{\infty} P(R, n), \quad (\text{B1})$$

where  $\gamma$  is a constant that we have varied between 3 and 10 and  $\text{floor}(x)$  is a function that returns the largest integer smaller than  $x$ . We find that  $P_p(R, \gamma)$  can distinguish the models considered in Figure B1 with comparable significance to  $P(R, 0)$ . Like  $P(R, 0)$ , a survey needs an estimate for its contamination fraction to be able to predict  $P_p(R, \gamma)$ . This can be estimated by performing spectroscopic follow-up on a portion of the survey. Unlike  $P(R, 0)$ , the bias in



**Figure C1.** Effect of the duty cycle of ionizing sources on the morphology of reionization. The white regions are ionized and the black are neutral. The left panels are from a simulation using model (i), and the right panels are from an identical simulation except that only 10% of the sources are active at any time, and these sources have 10 times the luminosity compared to the same sources in model (i). The active sources are chosen randomly, and this randomization is performed every 20 million years. The top panels have  $\bar{x}_i = 0.3$ , and the bottom panels have  $\bar{x}_i = 0.7$ . All panels are from a slice through the 186 Mpc box.

the estimation of  $P_p(R, \gamma)$  incurred by foreground contamination is not an exponential factor, and, therefore, a survey can more accurately correct for this bias with an estimate of  $F_c$ .

More work needs to be done to quantify how much independent information derives from void (and peak) statistics compared to the correlation function and to quantify the benefits of these statistics over the correlation function.

## APPENDIX C: DUTY CYCLE

In McQuinn et al. (2006), the effect of the morphology of reionization on the source properties, on the redshift of reionization, and on the minihalos were investigated in detail. McQuinn et al. (2006) found that the structure of the HII regions during reionization was robust to most considered effects, with the most important dependence being the value of the ionized fraction. However, the effect of a duty cycle for the ionizing sources – the fraction of time the sources are emitting ionizing photons – on the structure of the HII regions was omitted in their analysis. It is probable that star formation at high redshifts is sporadic. At any given time, a small fraction of galaxies may contain the high mass stars that produce ionizing photons. These active galaxies are probably also the LAEs.

It is interesting to understand the effect of a small ionizing photon duty cycle  $f_{\text{ion}} \ll 1$  on the structure of the HII

regions. None of the simulations used in the body of this paper included this effect. Figure C1 compares the 130 Mpc/h simulation using model (i) in which  $f_{\text{ion}} = 1$  (*left panels*) to a simulation where  $f_{\text{ion}} = 0.1$  and in which the sources that are active have a luminosity that is boosted by a factor of 10 over the sources in model (i) (*right panels*). To achieve  $f_{\text{ion}} = 0.1$ , 10% of the sources are randomly selected to be active, and this randomization is performed every 20 million years. Twenty-million years was chosen to roughly match the lifetime of massive stars. One can see in Figure C1 that the large HII regions are very similar between these two cases. This invariance owes to thousands of emitters in each large bubble contributing to its ionization, such that the total number of photons produced inside the bubble, and, therefore, the bubble size is largely unaffected by  $f_{\text{ion}}$ . A similar conclusion about the effect of  $f_{\text{ion}}$  was reached in the analytic study of Furlanetto et al. (2005).

The duty cycle does influence the smallest HII regions – bubbles small enough such that shot noise in the number of galaxies is important. The criterion for  $f_{\text{ion}}$  to change the LAE statistics from models in which all galaxies are active is that the fluctuations owing to shot noise in the total number of source galaxies must be comparable to the cosmological fluctuations in the number of these galaxies within bubbles of size  $\gtrsim 1$  pMpc. At  $z \sim 7$  there are  $\sim 10^4 f_{\text{ion}}$  galaxies emitting ionizing photons in an HII region of size 1 pMpc in the source models discussed in this paper, whereas cosmological fluctuations in the source abundance are of the order of unity at this scale. Therefore, it would require either  $f_{\text{ion}} \sim 10^{-3}$  for the value of  $f_{\text{ion}}$  to affect the statistics of the LAEs.

## APPENDIX D: SOURCE GROUPING

The Sokasian et al. (2001, 2003) and McQuinn et al. (2006) radiative transfer code has been improved for this paper to group sources more efficiently, speeding up the computation. In previous versions of the code, only sources that fall within the same grid cell were grouped as a single source, often resulting in tens of millions of sources that must be processed in a time step. This algorithm did not take advantage of the fact that the HII regions become much larger than the size of a grid cell, motivating a more aggressive grouping algorithm without a significant loss in accuracy.

Our algorithm for grouping sources is as follows:

(i) Smooth ionization field  $x_i$  at scale  $R$  with a top hat filter, yielding the field  $\tilde{x}_{i,R}$ .

(ii) Loop over sources in random order.

If a source is in a cell in which  $\tilde{x}_{i,R} > \zeta$ , check that  $\tilde{x}_{i,r} > \zeta$  is also satisfied at all  $r < R$ . If yes, group sources within a sphere of radius  $\eta R$  into a single source, and place the new source at the center of luminosity of the grouped sources.

(iii) Repeat previous steps for smaller  $R$ , but only using the sources which have not been grouped. Stop when  $\eta R$  is less than the width of a grid cell.

(iv) Take the new grouped source field and repeat all previous steps once again.

In this paper, we start smoothing with  $R = 50$  Mpc, and set  $\zeta = 0.9$  and  $\eta = 0.1$ . As a test of this algorithm, we have

cross correlated the ionization field generated from a simulation that has source grouping to one without source grouping, but with the same sources and at fixed  $\tilde{x}_i$ . We find no appreciable difference between the two ionization fields, implying that source grouping does not affect our conclusions. Source grouping speeds up the radiative transfer algorithm considerably when  $\tilde{x}_i \gtrsim 0.5$ .

More aggressive grouping does lead to less accuracy in determining the local photo-ionization rate around sources, which can influence the transmission properties of the Ly $\alpha$  line. However, we only use this more aggressive grouping algorithm in combination with Method 2 in Section 2 to calculate the line flux, and Method 2 does not depend on the details of the photo-ionization state of the gas.